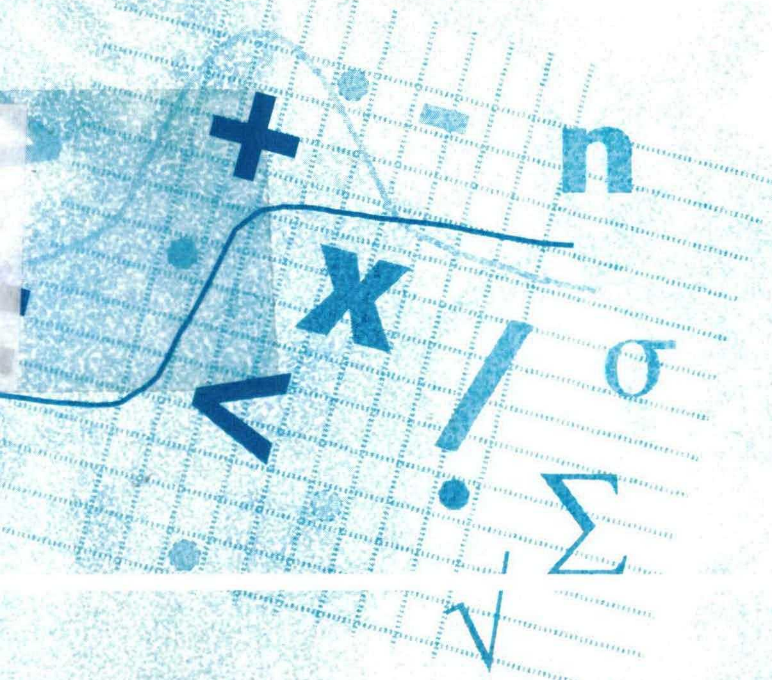
Tilburg University Press

Maximum Likelihood Estimation of Exact ARMA Models

Jan van der Leeuw



Stellingen

behorende bij het proefschrift

Maximum Likelihood Estimation of Exact ARMA Models

van

Jan van der Leeuw

18 juni 1997

1. Met de matrix identiteit $V^{-1}-V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} = Z(Z'VZ)^{-1}Z'$ met $Z'X=0$ en enkele vooronderstellingen betreffende dimensies en rang van voorkomende matrices kan eenvoudig worden aangetoond, dat de wegingsmatrix voor de fouten van een regressiemodel met MA(1) storingen eigenvectoren heeft, die onafhankelijk zijn van de waarde van de MA parameter. Voor MA(q) met q>1 geldt deze eigenschap niet.

2. Gezien de ruime belangstelling waarin tijdreeksanalyse zich heeft mogen verheugen, is het des te verwonderlijker dat aan de in dit proefschrift voorgestelde matrix benadering weinig aandacht is geschonken.

3. Een van de best bewaarde geheimen in tijdreeksanalyse is dat de door Box en Jenkins voorgestelde en alom bekende *back forecasting* methode zelden werkt.
G.E.P. Box en G.M. Jenkins, *Time Series Analysis: Forecasting and Control,* Revised Edition, Holden Day, San Francisco 1976, blz. 200.

4. De wijze waarop in de meeste handboeken tijdreeksanalyse en regressieanalyse worden gepresenteerd, wekt de indruk dat beide methoden elkaars tegenpolen zijn. Deze indruk is onjuist, ze vullen elkaar daarentegen aan.

5. Hoewel de degelijkheid van het werk van C.R. Rao onomstreden is, is ook hij soms slordig. In zijn bekende werk *Linear Statistical Inference* valt Problem 2.9, blz. 33, afgezien van een zetfout, op door het ontbreken van de voorwaarde dat de matrix zoals gepresenteerd niet-singulier dient te zijn.
C.R. Rao, *Linear Statistical Inference and its Applications,* Second Edition, John Wiley & Sons, New York, 1973.

6. Privatisering en marktwerking voor de sector sociale zekerheid zijn een *contradictio in terminis*: het gevolg zal zijn, dat de overheid na verloop van tijd de onverzekerde en onverzekerbare risico's voor haar rekening moet nemen.

7. In het licht van de Europese eenwording wordt de maatschappelijke en persoonlijke vorming eerder gediend door geschiedenis dan wiskunde in het voortgezet onderwijs verplicht te stellen.

8. De maatschappelijk noodzakelijke terugdringing van de automobiliteit wordt mede gediend door het imago van de autogebruiker te ondermijnen; lease-auto's van een opvallende kleur nummerbord voorzien is het overwegen waard.

9. Het niet belasten van vermogenswinsten werkt beleggen als piramidespel in de hand.

10. Ingeval *techno-lease* fiscaal acceptabel wordt geacht, dienen de mogelijkheden voor *brain-lease* nader bekeken te worden.

Maximum Likelihood Estimation of Exact ARMA Models

Maximum Likelihood Estimation of Exact ARMA Models

Proefschrift

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Brabant, op gezag
van de rector magnificus, prof.dr. L.F.W. de Klerk,
in het openbaar te verdedigen ten overstaan van
een door het college van decanen aangewezen commissie
in de aula van de Universiteit
op woensdag 18 juni 1997 om 14.15 uur

door

Johannes Leonardus van der Leeuw

geboren op 27 januari 1948 te Heerlen

Tilburg University Press 1997

Promotores: Prof. dr. B.B. van der Genugten
Prof. dr. J.R. Magnus
Copromotor: Dr. H.H. Tigelaar

ISBN 90-361-9867-4

PREFACE

Since the days of Gauss and Buys Ballot many papers and books have been written on the subject of time series analysis. Lots of ideas how to treat time series, elegantly formulated in matrix theory, can be found in numerous textbooks. However, a concise matrix approach was hardly present in the theory of autoregressive and moving average models. Only some minor results had been obtained and I wondered whether a closed matrix form for the covariance matrix of this type of models was possible. Initially, the question was interesting but as time progressed it became more or less obsessive.

Once having started to find a matrix formulation I discovered why it was lacking. It became a serious battle with little progress despite many numerical and theoretical experiments. Everywhere I went, I left a trace of paper sheets fully written with attempts to win the struggle against the equation that had to be solved. Patience, endurance and some luck brought the results I hoped for, but only after numerous decisions to stop the research and - indeed - to start it again.

Between the moment I found the solution and the completion of this study I got help and stimulation from many persons. I am grateful to many and a few of them I would like to mention here. First of all I am grateful to Ben van de Genugten, who noticed my work and offered me the opportunity to work for several years in the stimulating climate of Tilburg University; his critical attitude added greately to the final result. Jan Magnus was of great help: both his former and more recent publications gave much inspiration and his personal encouragement was very important. Special thanks is due to Harry Tigelaar who became both my teacher and friend. Through our profound discussions he taught me that one should treat mathematics very carefully and that formulating is an art in itself. Without his support theorems in this thesis would only have been acceptable for non-mathematicians - if there would have been a thesis.

Eventually I have to thank my greatest supporter, especially in those periods when the completion seemed a distant mirage: Joke.

April, 1997

CONTENTS

CONTENTS

I INTRODUCTION

1.1 The scope of this study

In 1982 T.W. Anderson and R.P. Mentz have written in a survey article on maximum likelihood estimation in autoregressive and moving average models: *A detailed study of the likelihood function ... will involve explicit expressions for the determinant and the components of the inverse of the covariance matrix. These are not available in the literature in closed form expressions for arbitrary finite orders.* Since then several procedures are developed to find the elements of the ARMA covariance matrix, but neither of them was in a closed form expression. The purpose of this study is to present such a closed form and to show how profitable such a form is.

We will start with some comments on the maximum likelihood (ML) approach, its properties and the way we will use this principle in our study. Next we will have a look on some recent studies in this field.
Subsequently (Chapter 2) we will present a closed form expression for the general ARMA covariance matrix, which of course includes the AR an MA covariance matrices as special cases. Using this closed form we can easily derive expressions for the first and second derivative of the likelihood function and the information matrix (Chapter 3 and 4).

Chapter 5 shows how the so called conditional least squares (CLS) approach to estimate ARMA parameters is related to the maximum likelihood method. Again, first and second order conditions are derived (Chapter 6). We conclude this study (Chapter 7) by presenting some simulation results. We will compare the estimation of several AR, MA and ARMA models using the maximum likelihood and the conditional least square method. As reported by other authors we conclude that ML estimators perform better.

1.2 The linear model with ARMA distributed errors

We will be concerned with the linear model

$$y = X\beta + \varepsilon, \tag{1.1}$$

where y is a (T×1) vector of observations, X a given (T×k) non-random matrix of rank k, β is a vector of length k and ε is a vector of length T. The estimation problem is to find estimates for β and for the variance of ε. In case ε is a vector of independently and identically distributed disturbances with variance σ^2, the estimator for β is usually based on the least square method, *i.e.* minimise the sum of squares of the elements of the vector $y-X\beta$ with respect to β. The result is the Ordinary Least Square estimator $b = (X^TX)^{-1}X^Ty$. This estimator has several pleasant properties. First, it is a linear function of y, secondly it is unbiased as $Eb=\beta$. Furthermore it is - within the class of linear and unbiased estimators - 'best', which means that its covariance matrix $E(b-\beta)(b-\beta)^T$ is minimal. The variance σ^2 is estimated by $e^Te/(T-k)$, where $e=y-Xb$.

Here however we suppose the components of the error vector ε to follow an ARMA scheme. The general form of an ARMA(p,q) error structure is given by

$$\varepsilon_t = -\sum_{i=1}^{p} \vartheta_i\varepsilon_{t-i} + v_t + \sum_{i=1}^{q} \alpha_i v_{t-i}, \quad t=0,\pm1,\ldots \tag{1.2}$$

where v_t $(t=0,\pm1,\ldots)$ is a sequence of independently and identically distributed random variables with mean zero and variance σ^2. We suppose p and q to be known. Both $(\vartheta_1..\vartheta_p)^T$ and $(\alpha_1..\alpha_q)^T$ are unknown parameters. This means that the error ε_t depends on the errors in the past $\varepsilon_{t-1},\ldots,\varepsilon_{t-p}$ and a random shock v_t and previous shocks $v_{t-1},\ldots,v_{t-q}$. Let ϑ denote the vector $(\vartheta_1,\vartheta_2,\ldots,\vartheta_p)^T$ of so called AR-parameters and α the vector $(\alpha_1,\alpha_2,\ldots,\alpha_q)^T$ of MA-parameters. We will use σ^2V to denote the covariance matrix of ε: $\sigma^2V = E\varepsilon\varepsilon^T$. In case $q=0$, $p>0$ we have an AR(p) model, in case $p=0$, $q>0$ we call it an MA(q) model. Essentially we will deal with the general ARMA(p,q) model in this study and consider the AR and MA model as special cases.

Now we not only have to estimate the vector β and σ^2, but also the unknown parameters ϑ and α. It is clear, that this model reduces to what is called a pure time series model in case X is zero: ε no longer unobservable, but identical to y.

We suppose that the matrix X is deterministic. The vector ε and thus the vector y are considered as a sample from a (discrete) stochastic process. A process z_t, $t=0,\pm1,\pm2,..$ is called weakly stationary or covariance stationary if its mean exists and the covariance between z_t and z_{t+k} is independent of t for all k (Judge *e.a.*, 1985). If we assume that the distribution of the set $\{z_1,z_2,\ldots,z_k\}$ is the same as for the set $\{z_{1+s},z_{2+s},\ldots,z_{k+s}\}$ for all k and s, we call such a process strictly stationary.

To insure that (1.2) has a unique weakly stationary solution of the form $\varepsilon_t = \sum_{k=0}^{\infty} c_k v_k$ the zeros of the so called associated polynomial of the AR part must all lie outside the unit circle (see *e.g.* T.W. Anderson, 1971, p.170):

$$f(z) = 1+\vartheta_1 z+\ldots+\vartheta_p z^p \neq 0 \text{ for } |z|\leq 1,\ z\in \mathbb{C}. \tag{1.3}$$

Sometimes this condition is called the stationarity condition (*e.g.* Box and Jenkins, 1976, p. 54). In Chapter 2 we will show that this condition is sufficient for positive definiteness of the covariance matrix of the disturbance vector. An MA process is always stationary as can easily be shown. Here the problem is that the representation of an MA process is not unique without restrictions. To guarantee a unique representation of an MA process, it is sufficient that the zeros of the associated polynomial lie on or outside the unit circle:

$$g(z) = 1+\alpha_1 z+\ldots+\alpha_q z^q \neq 0 \text{ for } |z|<1,\ z\in \mathbb{C}. \tag{1.4}$$

This condition is also known as the invertibility condition (Box and Jenkins, 1976, p.67). If this condition is satisfied the MA process can be rewritten as a one sided AR process.

1.3 Minimum Distance and Maximum Likelihood estimation

Two widely used methods for estimating unknown parameters are minimum distance and maximum likelihood. The differences between both methods are small, certainly when the number of observations is large. We will give a short introduction of both methods.

Suppose we have a linear model like (1.1), where the disturbances obey $E\varepsilon=0$ and $E(\varepsilon\varepsilon^T) = \sigma^2 V(\xi)$, indicating that the positive definite matrix V depends on a $(n \times 1)$ vector ξ of unknown parameters. As stated before we now have to estimate β, σ^2 and ξ. A numerical procedure to estimate β and ξ simultaneously is not available. The general method is to start from an initial value for ξ (and thus V) and to use this to estimate β. This estimate of β is used to find a new value for ξ and so on. We will discuss this in detail below.

1.3.1 Minimum Distance estimation

Following Malinvaud (Malinvaud, 1970, p. 325) we define minimum distance estimators as the vector b and the vector $\hat{\xi}$ for which the weighted sum of squares $S(\beta,\xi) = (y-X\beta)^T V^{-1}(\xi)(y-X\beta)$ is minimised. The reason to adopt this estimation method is as follows. When the errors are not independently distributed, the ordinary least squares estimator for β, given ξ, is indeed still unbiased, but not necessarily best. It can be shown that the best linear unbiased estimator can be found by minimising $S(\beta,\xi)$ with respect to β. This can be seen as follows.

To simplify expressions we now omit the argument ξ in $V(\xi)$. As V is a positive definite matrix, it can be written as the product of an inverted matrix and its transpose: $V = U^{-1}U^{-T}$, where we use the symbol $^{-T}$ for the inverse of a transposed matrix. Then the transformed model $Uy = UX + U\varepsilon$ has uncorrelated errors: $EU\varepsilon = UE\varepsilon = 0$ and $E(U\varepsilon)(U\varepsilon)^T = UE(\varepsilon\varepsilon^T)U^T = \sigma^2 I$. Applying the least squares rule on the transformed model, the sum of squares to be minimised is $(Uy-UX\beta)^T(Uy-UX\beta) = (y-X\beta)^T U^T U(y-X\beta) = (y-X\beta)^T V^{-1}(y-X\beta)$.

Differentiating to β and ξ gives:

$$\frac{\partial S}{\partial \beta} = -2(y-X\beta)^T V^{-1} X \tag{1.5}$$

and

$$\frac{\partial S}{\partial \xi_i} = (y-X\beta)^T \frac{\partial V^{-1}}{\partial \xi_i}(y-X\beta), \quad i=1,..,n \tag{1.6}$$

From (1.5) follows the so called Aitken estimator $b = (X^T V^{-1} X)^{-1} X^T V^{-1} y$ as estimator for β, if V is known. Since $\frac{\partial^2 S}{\partial \beta^2} = 2X^T V^{-1} X$ is positive definite, b minimises S for every choice of ξ. Define $e = y-Xb$ and it is clear that minimising $S(\beta,\xi)$ is equivalent to minimising

$$S_{MD}(\xi) = e^T V^{-1} e, \tag{1.7}$$

where the symbol S_{MD} is used to denote minimum distance estimation. Minimising $S_{MD}(\xi)$ gives ξ, which can be used to compute new values for β and ε. Notice that (1.7) amounts to solving $e^T\frac{\partial V^{-1}}{\partial \xi_i}e = 0 \ \forall \ i$, as $e^T V^{-1}\frac{\partial e}{\partial \xi_i} = = 2e^T V^{-1}\frac{\partial (y-Xb)}{\partial \xi_i} = -2(y-Xb)^T V^{-1} X \frac{\partial b}{\partial \xi_i} = -2(X^T V^{-1} y - X^T V^{-1} X b)^T \frac{\partial b}{\partial \xi_i} = 0$ in view of the definition of b.

Such a procedure will converge as demonstrated in a lemma by Oberdorfer and Kmenta (1974), if the parameter space is compact, but this is certainly not always the case as shown by Don and Magnus (1980).

1.3.2 Maximum Likelihood estimation

For ML estimation we assume that ε has a multivariate normal distribution with mean zero and covariance matrix $\sigma^2 V(\xi)$ with density p, given by $p(x) = (2\pi)^{-T/2} |\sigma^2 V|^{-1/2} \exp(-\frac{x^T V^{-1} x}{2\sigma^2})$. This assumption enables us to form the loglikelihood function:

$$L(\beta,\sigma^2,\xi \mid y) = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log \sigma^2 - \frac{1}{2}\log|V| - \frac{(y-X\beta)^T V^{-1}(y-X\beta)}{2\sigma^2}. \tag{1.8}$$

Maximum likelihood estimation seeks β, σ^2 and ξ such that L is maximal, given the dataset. The log is only taken to simplify the differentiation of the function.

Hence the derivatives are

$$\frac{\partial L}{\partial \sigma^2} = -\frac{T}{2}\frac{1}{\sigma^2} - \frac{(y-X\beta)^T V^{-1}(y-X\beta)}{2\sigma^4} \tag{1.9}$$

$$\frac{\partial L}{\partial \beta} = \frac{2X^T V^{-1}(y-X\beta)}{2\sigma^2} \tag{1.10}$$

$$\frac{\partial L}{\partial \xi_i} = -\frac{1}{2}\frac{\partial \log|V|}{\partial \xi_i} - (y-X\beta)^T\frac{\partial V^{-1}}{\partial \xi_i}(y-X\beta), \ i=1,..,n. \tag{1.11}$$

In Chapter 3 we will study the structure of (1.11) in detail. From (1.9) follows the ML-estimator $\frac{(y-X\beta)^T V^{-1}(y-X\beta)}{T}$ for σ^2 given V and β and from (1.10) $b=(X^T V^{-1} X)^{-1} X^T V^{-1} y$ as ML-estimator for β for given V. Substitute these values in (1.8), write e for y-Xb and after some manipulations we get $L(\xi) = -\frac{T}{2}\{1 + \log 2T\pi + \log|V|^{1/T}e^TV^{-1}e\}$ or

$L(\xi) = -\frac{T}{2}\{1 + \log 2T\pi\} - \frac{1}{2}(\log|V| + T\log e^TV^{-1}e)\}$.

This makes clear that minimising

$$S_{ML}(\xi) = |V|^{1/T}e^TV^{-1}e, \tag{1.12}$$

is equivalent to maximising the loglikelihood in order to find an estimate for ξ.

The only difference between S_{MD} (1.7) and S_{ML} (1.12) is the term $|V|^{1/T}$. The matrix V is of order T, but if the value of $|V|$ as function of T is bounded, then $|V|^{1/T}$ will approach 1 when T becomes large, as will be demonstrated in Chapter 5. This means that the values of S_{ML} and S_{MD} are almost equal. In many cases there is computationally hardly any difference between the two approaches.

Many authors have given derivations of ML estimators. For the most simple case, *viz.* the AR(1) model we refer to the much cited article of Beach and MacKinnon (1978). Magnus (1978) showed that it is possible to derive simultaneously maximum likelihood estimates of the regression parameters and the parameters of the covariance matrix with the iteration procedure outlined before and that such a procedure converges. Moreover he gives second derivatives of the likelihood function in a slightly different form and shows that maximum likelihood estimators are consistent, asymptotically normal and asymptotically efficient. Of course most textbooks give in one way or another a derivation of ML estimators, see *e.g.* Judge (1985).

1.3.3 Conditional Least Squares

Closely related to MD estimation is Conditional Least Squares (CLS) estimation. It differs from MD estimation in so far that the covariance matrix used in the function to be minimised does not correspond to the error specification as given in (1.2). In fact, the covariance matrix has a more simple structure under CLS estimation. This method is nowadays widely used for the estimation of ARMA models. The complete discussion will be found in Chapter 6. Eventually we will compare (exact) Maximum Likelihood with the Conditional Least Squares method.

1.4 Estimation of ARMA parameters

For both estimation methods, MD and ML, we need the inverse of the covariance matrix, for ML estimation moreover its determinant. Of course, if values for the elements of the covariance matrix are known, its inverse and determinant can be obtained numerically. However, as the covariance matrix has dimensions T by T, it is clear that numerical methods are rather inefficient for large T. This is one reason why analytical expressions, containing matrices of much smaller dimensions to be inverted are desirable. Another reason to use analytical expressions is the opportunity to investigate first order conditions and limiting forms when the number of observations is large. Therefore much interest has been shown in analytical and closed form formulas for the ARMA covariance matrix. But these were only available for specific cases like MA(q), AR(1) and AR(2).

The number of articles about the ARMA covariances is enormous. They can broadly be divided into three groups. The first one consists of contributions that describe the autocovariance function, which gives the individual covariances as a function of the ARMA parameters and a time parameter. The second group are algorithms which provide a computational tool to estimate the parameters from the (computed) covariances. The last category presents expressions in matrix form for the covariance matrix, but only for specific values of p and q. This study will give a general form in the next chapter. Now we will give some examples of what can be found in the ARMA literature.

Almost all studies start from an equation equal or similar to (1.2):
$\varepsilon_t + \vartheta_1 \varepsilon_{t-1} + .. + \vartheta_p \varepsilon_{t-p} = v_t + \alpha_1 v_{t-1} + .. + \alpha_q v_{t-q}, \ t = 0, \pm 1, ..$
This equation can be used in two ways. The first one is to regard the covariances $\gamma_k = E\varepsilon_t \varepsilon_{t-k}$ or the corresponding autocorrelations $\rho_k = \gamma_k / \gamma_0$ as given and to express the elements of α and ϑ explicitly or implicitly as function of the (auto)covariances. The second way is to find an expression for the theoretical covariances $E\varepsilon_t \varepsilon_{t-k}$ as function of α and ϑ. The values of α and ϑ have to be estimated by another method, *e.g.* Maximum Likelihood.

An example of the first way are the so called Yule-Walker equations. In case of a pure AR model an estimate for the parameters can be found in the following way. Multiplying both sides of (1.2) by ε_{t-k}, $k=1,..p$, and taking expectations gives a set of p equations, which contain the covariances, $\gamma_k = E\varepsilon_t\varepsilon_{t-k}$. After dividing by γ_0 one gets the so called autocorrelation function:

$$\rho_k = \vartheta_1\rho_{k-1} + \vartheta_2\rho_{k-2} + .. + \vartheta_k\rho_{k-p},\ k=1,..p. \tag{1.13}$$

Considering the vector ϑ as unknown it can be solved as function of the autocorrelations ρ_k, $k=1,..p$, which are estimated as $\sum_{i=k+1}^{T} \varepsilon_i\varepsilon_{i-k} / \sum_{i=1}^{T} \varepsilon_i^2$ (see *e.g.* Box and Jenkins, 1976, p. 54-55). For the MA model Box and Jenkins (1976, p.68) suggest to multiply ε_t $(= v_t + \alpha_1 v_{t1} + .. + \alpha_q v_{t-q})$ by ε_{t-k}, which gives after taking expectations and dividing by γ_0:

$$\rho_k = \begin{cases} \dfrac{-\alpha_k + \alpha_1\alpha_{k+1} + .. + \alpha_{q-k}\alpha_q}{1 + \alpha_1^2 + .. + \alpha_q^2} & \text{for } k=1,2,..,q \\ 0 & \text{for } k>q \end{cases} . \tag{1.14}$$

As they state, it is quite complicated to solve the vector α from these equations.

For the mixed process Box and Jenkins present (1976, p. 74-75) the difference equation

$$\gamma_k = -\vartheta_1\gamma_{k-1} - \vartheta_2\gamma_{k-2} - .. - \vartheta_p\gamma_{k-p} + \gamma_{ev}(k) + \alpha_1\gamma_{ev}(k-1) + .. + \alpha_q\gamma_{ev}(k-q) \tag{1.15}$$

where $\gamma_{ev}(k) = E\varepsilon_{t-k}v_t$, which is equal to zero if $k>0$, but not if $k\leq 0$. Box and Jenkins propose to compute these values by using a method which they called *back forecasting* (1976, p. 199-200). Although it is not stated explicitly (1.15) is to present the autocorrelation function, not a way to estimate the elements of α and ϑ.

McLeod (1975, 1977) noticed that multiplying (1.2) by v_{t-k} and taking expectations gives

$Ev_{t-k}\varepsilon_t + \vartheta_1 Ev_{t-k}\varepsilon_{t-1} + .. + \vartheta_p Ev_{t-k}\varepsilon_{t-p} = Ev_{t-k}v_t + \alpha_1 Ev_{t-k}v_{t-1} + .. + \alpha_i Ev_{t-k}v_{t-i}$ or

$$\gamma_{ev}(-k) + \gamma_{ev}(-k+1) + .. + \gamma_{ev}(-k+p) = \begin{cases} \alpha_k\sigma^2 \text{ for } k=0,1,..,q, \\ 0 \text{ elsewhere} \end{cases} \tag{1.16}$$

with $\alpha_0 = 1$. Using (1.15) and (1.16) he develops an algorithm to solve for γ_k. Tunnicliffe Wilson (1979) uses this set of equations to do the opposite and finds a recursive scheme to compute α and ϑ from the covariances.

Other authors like Nerlove *e.a.* (p. 78-85) obtain expressions for the covariances using the Laurent series expansion of the so-called autocovariance generating function by applying Cauchy's integral formula and the residue theorem. He arrives at the formula

$$\gamma(k) = \sigma^2 \sum_{i=1}^{p} \frac{\vartheta_i^{p+|k|-q-1} \prod_{j=1}^{q} (1-\alpha_j\vartheta_i)(\vartheta_i-\alpha_j)}{\prod_{j=1}^{p} (1-\vartheta_j\vartheta_i) \prod_{j=1,j\neq i}^{p} (\vartheta_i-\vartheta_j)} \tag{1.17}$$

which holds when no multiple roots occur (Nerlove, 1979, p. 79-80).

From a theoretical point of view the above mentioned methods to express the covariances in α and ϑ are sufficient to compute the desired estimators. From the pair equations (1.15) and (1.16) or from (1.17) one can compute the covariances, which are needed to form the covariance matrix. Next compute b and e, and find new values for α and ϑ by minimising (1.7) or (1.12). The problems are the computation of the inverse and the determinant of the covariance matrix which take a lot of computer time when T becomes large.

Moreover using algorithms can be inaccurate because of round off errors. This can be avoided by using analytical expressions for the covariances. However these become very complicated, except for simple cases like low order AR and MA cases. Even for the ARMA(1,1) case formulas are long and tedious as shown by Tiao and Ali (1971). In section (2.5) we present some examples.

As the inverse and the determinant of the covariance matrix are needed to obtain estimates, several authors proposed methods to find them. One example is the article of Galbraith and Galbraith (1974) who present a matrix expression for the inverse and determinant of the covariance matrix. They obtain expressions for the inverse in the form of the product of several matrices. However the elements of these matrices are not simple expressions of the parameters, but have to be computed using an algorithm. Remarkable is their derivation of the formula of the AR(p) covariance matrix in closed form, but they did not notice this. Also De Gooijer (1978) gives a formula for the inverse, be it that one of the components of his matrix expression is a covariance matrix of low order that has to be found and computed separately. A general method to find this matrix is not presented. An author who claims to present an exact formula for the inverse of the covariance matrix is Zinde-Walsh (1988, 1990). She observes that the inverse process of an AR process is similar to an MA process and vice versa. The elements of the inverse infinite process, as she calls it, are obtained in a way similar to (1.17) as given by Nerlove. Zinde-Walsh gives the difference between the exact inverse and the infinite inverse in the form of matrices. As these forms are rather intractable she presents also approximations. However the forms are still so complicated that analytical differentiation is impossible. Furthermore the structures of the covariance matrix and its inverse are not clear.

Another approach is to use a matrix which transforms the data in such a way, that a straightforward OLS computation gives the Aitken estimator for β. As the covariance matrix is positive definite it can be expressed as the product of a lower triangular matrix and its transpose. Such a transformation matrix can be found numerically using a Cholesky decomposition. Analytical expressions for the transformation matrix, if available, are again quite complicated. An approach like this can be found at Knottnerus (1989). He presents a two step procedure, where in the first step starting values for γ_t, $t=1,..,\max(p,q)$ are computed which are used in the second step to find the remaining covariances. The latter is most simple. From (1.15) follows for $k>0$ $\gamma_k = -\vartheta_1\gamma_{k-1}-\vartheta_2\gamma_{k-2}-..-\vartheta_p\gamma_{k-p}$, which gives a recursive formula for γ_k. The former step is quite complicated. Knottnerus shows that it is possible to write $(\gamma_0\ \gamma_1..\gamma_{\max(p,q)})^T$ explicitly as a function of α and ϑ. Furthermore he gives an algorithm which gives a lower triangular matrix that transforms the data such that the errors are independently and identically distributed. His derivations are lengthy and complicated. No formula is given for the inverse of the covariance matrix. Only F.X. Diebold (1986) found a closed form expression for the MA(q) case, which as will become clear, is the most simple one. In the next chapter we will present an alternative which does not suffer from the shortcomings mentioned here.

II THE ARMA COVARIANCE MATRIX IN CLOSED FORM

2.1 Introduction

As stated in the previous chapter a simple closed form formula will now be presented. In an article called *the treatment of autocorrelation* Pagan (1974) proposes to write (1.2) in matrix form, but he gives the honour of the idea to A.W. Phillips, who should have mentioned this in a paper presented at the *Meeting of the Econometric Society* already in 1966. Using this matrix form for the parameters of the MA and AR part it is straightforward to get an expression for the covariance matrix of the error vector. The problem, however, is that the covariance matrix appears at both sides of the equal sign, which means that a matrix equation has to be solved. This is possible if the stationarity condition is fulfilled. The result is rather simple, but the way to find the solution is far from obvious.

2.2 Matrix form for ARMA parameters

Our starting point is the ARMA(p,q) error structure as given in (1.2):

$\varepsilon_t = -\sum_{i=1}^{p} \vartheta_i \varepsilon_{t-i} + v_t + \sum_{i=1}^{q} \alpha_i v_{t-i}$, $t=0, \pm1,..,$where v_t is a white noise: $Ev_t=0$, $Ev_t^2=\sigma^2$, $Ev_tv_s=0$ for $t\neq s$. As before we assume that the usual stationarity (1.3) and invertibility (1.4) conditions hold. We assume that T $(\geq p+q+1)$ observations are available.

The aim of this section is to rewrite (1.2) in matrix form. To do so we define a special type of Toeplitz matrices, namely a lower band matrix, say $B_{n,m}(\phi) \in \mathbb{R}^{nxm}$, with $\phi \in \mathbb{R}^n$, with elements $b_{i,j}$:

$$b_{i,j} = \begin{cases} \phi_{i-j} & \text{for } 1\leq j\leq i\leq n \\ 0 & \text{for } 1\leq i\leq j-1 \end{cases}.$$

In case $B_{n,m}(\phi)$ is square, its inverse of can be obtained by a simple algorithm. Another important characteristic of square Toeplitz matrices is the property that their product is a matrix of the same type and moreover that pairs of them commute: $B_{n,n}(\phi)B_{n,n}(\varphi) = B_{n,n}(\varphi)B_{n,n}(\phi) = B_{n,n}(\xi)$. Occasionally we need a (n×m) matrix consisting of only zeros, which we will write as $O_{n,m}$.

Following Pagan (1974), we introduce two matrices for both the AR parameters and the MA parameters. For the AR part we define a square lower band matrix $P = B_{T,T}(1,\vartheta_1,..,\vartheta_p,0,..0) \in \mathbb{R}^{T\times T}$ and a upper band matrix $Q = B^{T}_{p,T}(\vartheta_p,..,\vartheta_1) \in \mathbb{R}^{T\times p}$. When $p=0$, we get $P = B_{T,T}(1,0,..0) = I_T$, the unit matrix, while Q is not defined. It will become clear that it is useful to partition P and Q after p rows and columns:

$$P = \left[\begin{array}{c|c} P_{11} & O_{p,T-p} \\ \hline P_{12} & P_{22} \end{array}\right] = \left[\begin{array}{cccc|ccccc}
1 & & & & & & & & \\
\vartheta_1 & 1 & & & & & & & \\
 & \vartheta_1 & \ddots & & & & & & \\
\vdots & & \ddots & & & & & & \\
\vartheta_{p-1} & & \cdots & 1 & & & & & \\
\hline
\vartheta_p & \vartheta_{p-1} & \cdots & \vartheta_1 & 1 & & & & \\
 & \ddots & & \vdots & \vartheta_1 & \ddots & & & \\
 & & \ddots & & & \ddots & \ddots & & \\
 & & & \vartheta_{p-1} & & & \ddots & & \\
 & & & \vartheta_p & \vartheta_{p-1} & & & & \\
 & & & & \vartheta_p & \ddots & & & \\
 & & & & & \ddots & \ddots & & \\
 & & & & & \ddots & \ddots & 1 & \\
 & & & & & \vartheta_p & \vartheta_{p-1} \cdots & \vartheta_1 & 1
\end{array}\right].$$

Observe that $P_{11} = B_{p,p}(1,\vartheta_1,..\vartheta_{p-1})$, $P_{12} = B^{T}_{T-p,p}(\vartheta_p,..,\vartheta_1)$ and $P_{22} = B_{T-p,T-p}(1,\vartheta_1,..,\vartheta_p,0,..0)$. As we will encounter P_{11} frequently in the sequel we will use $\underline{P}$ for P_{11} to avoid the use of subscripts.

In a similar way we have

$$Q = \left[\begin{array}{c} Q_1 \\ \hline 0_{T-p,p} \end{array}\right] = \left[\begin{array}{cccc} \vartheta_p & \vartheta_{p-1} & \cdot & \vartheta_1 \\ 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & & \\ 0 & & & \vartheta_p \\ \hline 0 & & & 0 \\ \cdot & & & \cdot \\ & & & \\ 0 & \cdot & & 0 \end{array}\right], \text{ where } Q_1 = B^T_{p,p}(\vartheta_p,..,\vartheta_1) = \underline{Q}.$$

For the MA parameters we define in the same way $M = B_{T,T}(1,\alpha_1,..,\alpha_q,0,..0) \in \mathbb{R}^{T\times T}$ and $N = B^T_{q,T}(\alpha_q,..,\alpha_1) \in \mathbb{R}^{T\times q}$:

$$M = \left[\begin{array}{c|c} M_{11} & 0_{p,T-p} \\ \hline M_{12} & M_{22} \end{array}\right] = \left[\begin{array}{cccccc|cccccc} 1 & & & & & & & & & & & \\ \alpha_1 & 1 & & & & & & & & & & \\ & \alpha_1 & \cdot & & & & & & & & & \\ \cdot & & \cdot & & & & & & & & & \\ & & & & \vdots & & & & & & & \\ \alpha_{q-1} & & & & \cdot & 1 & & & & & & \\ \hline \alpha_q & \alpha_{q-1} & & & & \alpha_1 & 1 & & & & & \\ & \cdot & \cdot & & & \cdot & \alpha_1 & \cdot & & & & \\ & & & & & & & \cdot & \cdot & & & \\ & & & \cdot & \cdot & & & & \cdot & & & \\ & & & & \cdot & \alpha_{q-1} & & & & & & \\ & & & & & \alpha_q & \alpha_{q-1} & & & & & \\ & & & & & & \alpha_q & \cdot & & & & \\ & & & & & & & \cdot & \cdot & & & \\ & & & & & & & & \cdot & \cdot & \cdot & 1 \\ & & & & & & & & \alpha_q & \alpha_{q-1} & \alpha_1 & 1 \end{array}\right].$$

Again $M_{11} = B_{q,q}(1,\alpha_1,..\alpha_{q-1}) = \underline{M}$, $M_{12} = B^T_{T-q,q}(\alpha_q,..,\alpha_1)$ and $M_{22} = B_{T-q,T-q}(1,\alpha_1,..,\alpha_q,0,..0)$.

$$N = \left[\begin{array}{c} N_1 \\ \hline 0_{T-q,q} \end{array}\right] = \left[\begin{array}{cccc} \alpha_q & \alpha_{q-1} & \cdot & \alpha_1 \\ 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & & \\ 0 & & & \alpha_q \\ \hline 0 & & & 0 \\ \cdot & & & \cdot \\ & & & \\ 0 & \cdot & & 0 \end{array}\right], \text{ where } N_1 = B^T_{q,q}(\alpha_q,..,\alpha_1) = \underline{N}.$$

The partitioning of M and N is after q rows and columns. Where appropriate we partition after max(p,q) rows; this gives no loss of generality for two reasons. First it is always possible to fill the shorter vector of parameters with zeros. Secondly it is not essential to partition after exact p, respectively q rows, but p, respectively q, rows is minimal number for useful partitioning. It is clear that the determinant of P and M is equal to 1.

Next define $\varepsilon=(\varepsilon_1,..,\varepsilon_T)^T$, $v=(v_1,..v_T)^T$ and the auxiliary vectors $\bar{\varepsilon}$ and $\bar{v}$:

$\bar{\varepsilon}=\ (\varepsilon_{-p+1},\varepsilon_{-p+2},\ldots,\varepsilon_{-1},\varepsilon_0)^T,$

$\bar{v}=\ (v_{-q+1},v_{-q+2},\ldots,v_{-1},v_0)^T.$

Then we can write (1.2) in matrix form:

$$\begin{bmatrix} Q & P \end{bmatrix}\begin{bmatrix} \bar{\varepsilon} \\ \varepsilon \end{bmatrix} = \begin{bmatrix} N & M \end{bmatrix}\begin{bmatrix} \bar{v} \\ v \end{bmatrix}. \tag{2.1}$$

To relate the stationarity condition to these matrices we give the following theorem:

Theorem 2.1

Let $\underline{P}$ and $\underline{Q}$ be defined as above. The stationarity condition is equivalent to the condition that all solutions to $|\lambda\underline{P}+\underline{Q}|=0$ satisfy $|\lambda|<1$.

Proof

Denote $z_k=\ \lambda^{1/p}e^{2ki\pi/p}$, $k=1,2,..,p$, and as before let f be the polynomial associated with the AR part, *i.e.* $f(z)=\sum_{k=0}^{p}\vartheta_k z^k$. Since

$$\lambda\underline{P}+\underline{Q}=\begin{bmatrix} \lambda+\vartheta_p & \vartheta_{p-1} & \cdot & \vartheta_1 \\ \lambda\vartheta_1 & \lambda+\vartheta_p & \cdot & \vartheta_2 \\ \cdot & \cdot & & \cdot \\ \lambda\vartheta_{p-1} & \lambda\vartheta_{p-2} & \cdot & \lambda+\vartheta_p \end{bmatrix}$$

and $z_k^p=\lambda$, we see by direct verification that $\lambda\underline{P}+\underline{Q}$ has p linearly independent eigenvectors $(1\ z_k\ .\ .\ .\ z_k^{p-1})^T$, $k=1,..,p$ with corresponding eigenvalues $\mu_k=z_k^p f(1/z_k)$ (see *e.g.* Davies, 1979).

▸ Suppose that the stationarity condition is fulfilled. Let λ_0 be an arbitrarily solution of $|\lambda\underline{P}+\underline{Q}|=0$. Then $\lambda_0\neq 0$ since $\underline{Q}$ is non-singular and so $\lambda_0^{-1/p}e^{-2ki\pi/p}$ is a zero of f for some k lying outside the unit circle which implies $|\lambda_0|<1$.

▸ Conversely suppose that all solutions of $|\lambda\underline{P}+\underline{Q}|=0$ satisfy $|\lambda|<1$. Let z_0 be an arbitrary zero of f. Then $z_0\neq 0$, and since for $\lambda=z_0^{-p}$ the matrix $\lambda\underline{P}+\underline{Q}$

has eigenvalue $z_0^{-p}f(z_0)=0$ it follows that z_0^{-p} solves $|\lambda\underline{P}+\underline{Q}|=0$, and so we must have $|z_0^{-p}|\leq 1$ or equivalently $|z_0|>1$. □

Remark
If P and Q have dimensions r×r, while the number of AR parameters is $p < R$ then $|\lambda P+Q|=0$ has (r-p) solutions equal to zero. The equation $|\lambda P+Q|=0$ is equivalent to $|\lambda I_r+P^{-1}Q|=0$. As $\text{rank}(P^{-1}Q)=\text{rank}(Q)=p$, (r-p) eigenvalues have to be zero.

2.3 Covariance equation

Now we will derive an equation from which the exact covariance matrix can be solved. First we rewrite the error vector in matrix form. As done by several authors (De Gooijer, 1978 or Galbraith and Galbraith, 1974) we form an equation for the covariance matrix. But there is one difference as our equation involves only one unknown matrix. The solution of this covariance equation will be given in the next section.
For reason that become clear in the course of the proof we fill up the shorter vector of parameters by zeros. This will give no loss of generality as we will show. For the matrices P and M this has no consequences. If $p<q$ than the first q-p columns of Q are zeros, while N is unchanged. If $p>q$ the first p-q columns are N are zeros. Denoting the covariance matrix by σ^2V we state

Theorem 2.2
Suppose the stationarity condition holds. Then the matrix $V \in \mathbb{R}^{T\times T}$ corresponding to the ARMA(p,q) error specification is a solution of the equation

$$PVP^T = NN^T + MM^T + \begin{bmatrix} Q & 0 \end{bmatrix} V \begin{bmatrix} Q & 0 \end{bmatrix}^T - \begin{bmatrix} N & 0 \end{bmatrix} M^T P^{-T} \begin{bmatrix} Q & 0 \end{bmatrix}^T - \begin{bmatrix} Q & 0 \end{bmatrix} P^{-1} M \begin{bmatrix} N & 0 \end{bmatrix}^T \tag{2.2}$$

where $r=\max(p,q)$, $P=B_{T,T}(1,\vartheta_1,..,\vartheta_r,0,..0)$, $Q=B^T_{r,T}(\vartheta_r,..,\vartheta_1)$, $M=B_{T,T}(1,\alpha_1,..,\alpha_r,0,..0)$ and $N=B^T_{r,T}(\alpha_r,..,\alpha_1)$ and $0=0_{T,T-r}$.

Proof
From (2.1) we conclude $P\varepsilon = N\check{v}+Mv-Q\check{\varepsilon}$. To get the covariance matrix $E\varepsilon\varepsilon^T$ we postmultiply both sides by its transpose and take expectations, which

gives $PE(\varepsilon\varepsilon^T)P^T = E(N\dot{v}+Mv-Q\dot{\varepsilon})(N\dot{v}+Mv-Q\dot{\varepsilon})^T$. The right hand side contains the expressions $E\dot{v}\dot{v}^T$, $E\dot{v}v^T$, $E\dot{v}\dot{\varepsilon}^T$, Evv^T, $Ev\dot{\varepsilon}^T$ and $E\dot{\varepsilon}\dot{\varepsilon}^T$. Since v is vector of white noise, we have $Evv^T=\sigma^2 I_T$, $E\dot{v}\dot{v}^T=\sigma^2 I_r$ and $Ev\dot{v}^T=0$. Because we assume that the ARMA(p,q) process is stationary over time we have the same structure for $E\dot{\varepsilon}\dot{\varepsilon}^T$ as for $E\varepsilon\varepsilon^T$, *i.e.* $\sigma^2 V$. As the vector $\dot{\varepsilon}$ depends only on v_0, v_{-1}, ... (which are by assumption uncorrelated with v_1, v_2, ...), we conclude $Ev\dot{\varepsilon}=0$.
The resulting equation, which contains also $E\dot{\varepsilon}\dot{v}^T$, can be found in *e.g.* Galbraith and Galbraith, 1974 or de Gooijer, 1978. But we can go one step further, for the covariances of $\dot{\varepsilon}$ and $\dot{v}$ have - supposing stationarity - the same structure as the covariances of ε and v. This covariance can be derived as follows:
$E(P\varepsilon v^T) = E(N\dot{v}+Mv-Q\dot{\varepsilon})v^T = NE(\dot{v}v^T)+ME(vv^T)-QE(\dot{\varepsilon}v^T) = \sigma^2 M$,
which gives $E(\varepsilon v^T) = \sigma^2 P^{-1}M$. For $E(\dot{\varepsilon}\dot{v}^T)$ we get the first r rows and the first r columns of $P^{-1}M$. Here we need the modification of either Q or N to assure that the correct part of $P^{-1}M$ is used. Using $0 \in \mathbb{R}^{T\times(T-r)}$ for a matrix which consists of only zeros and dividing by σ^2 gives equation (2.2).□

2.4 Solution of the covariance equation

The problem of finding V is thus reduced to the problem of finding a solution of (2.2). We will show that this is possible if the stationarity condition holds.

Theorem 2.3
The covariance equation (2.2) has a unique solution if the stationarity condition fulfilled. The solution is given by
$$V = [N\ M](\bar{P}^T\bar{P}-\bar{Q}\bar{Q}^T)^{-1}[N\ M]^T, \tag{2.3}$$
where $M=B_{T,T}(1,\alpha_1,..,\alpha_q,0,..0)$, $N=B^T_{q,T}(\alpha_q,..,\alpha_1)$,
$\bar{P}=B_{T+q,T+q}(1,\vartheta_1,..,\vartheta_p,0,..0)$ and $\bar{Q}=B^T_{p,T+q}(\vartheta_p,..,\vartheta_1)$.

The proof is given by direct verification, but in stead of (2.3) an alternative expression for the covariance matrix will be used. Before we give the proof of Theorem 2.3 we give two lemmas: one for the existence of the inverted part and one for an alternative expression for (2.3).

Lemma 2.1
$\underline{P}^T\underline{P}\text{-}\underline{Q}\underline{Q}^T$ is not singular if the stationarity condition holds.

Proof

First observe that $\underline{P}^T\underline{P}\text{-}\underline{Q}\underline{Q}^T = \underline{P}\underline{P}^T\text{-}\underline{Q}^T\underline{Q}$. The matrices $\begin{bmatrix} \underline{P} & 0 \\ \underline{Q} & \underline{P} \end{bmatrix}$ and $\begin{bmatrix} \underline{Q}^T & 0 \\ \underline{P}^T & \underline{Q}^T \end{bmatrix}$ are both square lower band matrices. As they commute, we have $\underline{Q}\underline{Q}^T + \underline{P}\underline{P}^T = \underline{P}^T\underline{P} + \underline{Q}^T\underline{Q}$ or $\underline{P}^T\underline{P}\text{-}\underline{Q}\underline{Q}^T = \underline{P}\underline{P}^T\text{-}\underline{Q}^T\underline{Q}$.
From (2.2) we can form the matrix equation $\underline{P}\Delta\underline{P}^T = I_r + \underline{Q}\Delta\underline{Q}^T$, where all matrices are of order $r{\times}r$, $r{\geq}p$. It is the upper left part of (2.2) in the pure AR case. As $\underline{P}$ is non-singular we can substitute the left hand side in the right hand side:
$\underline{P}\Delta\underline{P}^T = I_r + \underline{Q}\underline{P}^{-1}(\underline{P}\Delta\underline{P}^T)\underline{P}^{-T}\underline{Q}^T$.
Repeated substitution gives, writing A for $\underline{Q}\underline{P}^{-1}$,
$\underline{P}\Delta\underline{P}^T = I + AA^T + A^2(A^T)^2 + A^3(A^T)^3 + \ldots$.
In view of theorem 2.1 the eigenvalues of A are less than one in absolute value or $\lim_{n\to\infty} A^n = 0$. As the right hand side consists of positive definite matrices, $\underline{P}\Delta\underline{P}^T$ and thus Δ has to be non-singular. This permits to take inverses at both sides of the equation: $(\underline{P}\Delta\underline{P}^T)^{-1} = (I + \underline{Q}\Delta\underline{Q}^T)^{-1}$ or, using an expression for the inverse of the sum of two matrices (see *e.g.* Rao, 1973):
$\underline{P}^{-T}\Delta^{-1}\underline{P}^{-1} = I\text{-}\underline{Q}(\underline{Q}^T\underline{Q} + \Delta^{-1})^{-1}\underline{Q}^T$.
Direct verification shows that $\Delta^{-1} = \underline{P}^T\underline{P}\text{-}\underline{Q}\underline{Q}^T = \underline{P}\underline{P}^T\text{-}\underline{Q}^T\underline{Q}$ is the solution of the last equation:
$\underline{P}^{-T}(\underline{P}^T\underline{P}\text{-}\underline{Q}\underline{Q}^T)\underline{P}^{-1} = I\text{-}\underline{Q}(\underline{Q}^T\underline{Q} + \underline{P}\underline{P}^T\text{-}\underline{Q}^T\underline{Q})^{-1}\underline{Q}^T$
or
$I\text{-}\underline{P}^{-T}\underline{Q}\underline{Q}^T\underline{P}^{-1} = I\text{-}\underline{Q}(\underline{P}\underline{P}^T)^{-1}\underline{Q}^T$.
Both sides are equal as $\underline{P}^{-1}$ and $\underline{Q}^T$ commute. Hence $(\underline{P}^T\underline{P}\text{-}\underline{Q}\underline{Q}^T)^{-1}$ exist. □

The next lemma gives an alternative expression for (2.3) and will be substituted into the covariance equation.

Lemma 2.2
Expression (2.3) is equivalent to
$$V = P^{-1}[MM^T + (PN\text{-}MQ)\underline{D}^{-1}(PN\text{-}MQ)^T]P^{-T}, \qquad (2.4)$$
where $\underline{D} = \underline{P}^T\underline{P}\text{-}\underline{Q}\underline{Q}^T \in \mathbb{R}^{r\times r}$.

Proof

We partition $\bar{P}$ and $\bar{Q}$ after r rows and columns and get

$$\bar{P}^T\bar{P}-\bar{Q}\bar{Q}^T = \begin{bmatrix} \underline{P}^T & P_{12}^T \\ 0 & P_{22}^T \end{bmatrix}\begin{bmatrix} \underline{P} & 0 \\ P_{12} & P_{22} \end{bmatrix} - \begin{bmatrix} \underline{Q} \\ 0 \end{bmatrix}\begin{bmatrix} \underline{Q}^T & 0 \end{bmatrix} = \begin{bmatrix} \underline{PP}^T & Q^TP \\ P^TQ & P^TP \end{bmatrix}$$

because $P_{22}=P \in \mathbb{R}^{T\times T}$, $P_{12}=Q \in \mathbb{R}^{T\times r}$ and $\underline{P}^T\underline{P}+P_{12}^TP_{12}-\underline{QQ}^T = \underline{PP}^T \in \mathbb{R}^{r\times r}$. As is easily verified, the inverse of this partitioned matrix is $\begin{bmatrix} \underline{D}^{-1} & -\underline{D}^{-1}Q^TP^{-T} \\ -P^{-1}Q\underline{D}^{-1} & P^{-1}P^{-T}+P^{-1}Q\underline{D}^{-1}Q^TP^{-T} \end{bmatrix}$. Premultiplying by [N M] and postmultiplying by its transpose gives (2.4), because P (and thus P^{-1}) and M commute. The existence of $\underline{D}^{-1}$ follows from the previous lemma. □

Proof of theorem 2.3

First we show that the covariance equation has a unique solution. Writing (2.2) in *vec*-notation and rearranging terms we see that uniqueness is guaranteed if $P\otimes P-[Q\ 0]\otimes[Q\ 0]$ is not singular. Its determinant is:

$$\begin{aligned} &|P\otimes P-[Q\ 0]\otimes[Q\ 0]| = \\ &\quad = |I-[Q\ 0]\otimes[Q\ 0][P\otimes P]^{-1}|\,|P\otimes P| \\ &\quad = |I-[Q\ 0]P^{-1}\otimes[Q\ 0]P^{-1}|. \end{aligned}$$

Hence, a sufficient condition for non-singularity is that all eigenvalues of $[Q\ 0]P^{-1}$ are less than one in absolute value. These eigenvalues are zero or equal to those of $\underline{QP}^{-1}$. As Theorem 2.1 states that $|\lambda|$ is less than one when the stationarity condition holds, we conclude that the determinant is nonzero, and thus that (2.2) has a unique solution. In case $p<q$ the rank of $\underline{Q}$ is at most p and we have q-p zero eigenvalues for $\underline{QP}^{-1}$.

Expression (2.4) contains the (T×r) matrix PN-MQ, which has only non-zero elements in the upper (r×r) part:

$$PN-MQ = \begin{bmatrix} \underline{P} & 0 \\ P_{12} & P_{22} \end{bmatrix}\begin{bmatrix} \underline{N} \\ 0 \end{bmatrix} - \begin{bmatrix} \underline{M} & 0 \\ M_{12} & M_3 \end{bmatrix}\begin{bmatrix} \underline{Q} \\ 0 \end{bmatrix} = \begin{bmatrix} \underline{PN}-\underline{MQ} \\ P_{12}\underline{N}-M_{12}\underline{Q} \end{bmatrix} = \begin{bmatrix} \underline{PN}-\underline{MQ} \\ 0 \end{bmatrix}.$$

Here $P_{12}\underline{N}-M_{12}\underline{Q} = \begin{bmatrix} \underline{Q} \\ 0 \end{bmatrix}\underline{N}-\begin{bmatrix} \underline{N} \\ 0 \end{bmatrix}\underline{Q} = 0$, because $\underline{Q}$ and $\underline{N}$ commute. Furthermore

$\underline{PN}-\underline{MQ} = \underline{NP}-\underline{QM}$, as $\begin{bmatrix} \underline{Q} & 0 \\ \underline{P} & \underline{Q} \end{bmatrix}$ and $\begin{bmatrix} \underline{N} & 0 \\ \underline{M} & \underline{N} \end{bmatrix}$ commute.

Now we prove that (2.4) a solution of (2.2). Substitute (2.4) in the left hand side of (2.2):

$$\begin{aligned} PVP^T &= P(P^{-1}[MM^T + (PN\text{-}MQ)\underline{D}^{-1}(PN\text{-}MQ)^T]P^{-T})P^T \\ &= MM^T + (PN\text{-}MQ)\underline{D}^{-1}(PN\text{-}MQ)^T \\ &= MM^T + \begin{bmatrix} (\underline{PN}\text{-}\underline{MQ})\underline{D}^{-1}(\underline{PN}\text{-}\underline{MQ})^T & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

The right hand side of (2.2) is

$$NN^T + MM^T + \begin{bmatrix} Q & 0 \end{bmatrix} V \begin{bmatrix} Q & 0 \end{bmatrix}^T - \begin{bmatrix} N & 0 \end{bmatrix} M^T P^{-T} \begin{bmatrix} Q & 0 \end{bmatrix}^T - \begin{bmatrix} Q & 0 \end{bmatrix} P^{-1} M \begin{bmatrix} N & 0 \end{bmatrix}^T.$$

Here we have :

$$NN^T = \begin{bmatrix} \underline{N} \\ 0 \end{bmatrix} \begin{bmatrix} \underline{N}^T & 0 \end{bmatrix} = \begin{bmatrix} \underline{NN}^T & 0 \\ 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} Q & 0 \end{bmatrix} V \begin{bmatrix} Q & 0 \end{bmatrix}^T = \begin{bmatrix} \underline{Q} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \underline{V} & .. \\ .. & .. \end{bmatrix} \begin{bmatrix} \underline{Q}^T & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \underline{QVQ}^T & 0 \\ 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} N & 0 \end{bmatrix} M^T P^{-T} \begin{bmatrix} Q & 0 \end{bmatrix}^T = \begin{bmatrix} \underline{N} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \underline{MP}^{-T} & .. \\ .. & .. \end{bmatrix} \begin{bmatrix} \underline{Q}^T & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \underline{NMP}^{-T}\underline{Q}^T & 0 \\ 0 & 0 \end{bmatrix}$$

and

$$\begin{bmatrix} Q & 0 \end{bmatrix} P^{-1} M \begin{bmatrix} N & 0 \end{bmatrix}^T = \begin{bmatrix} \underline{QP}^{-1}\underline{MN}^T & 0 \\ 0 & 0 \end{bmatrix}.$$

We conclude that, apart of MM^T, all parts except the upper left part, are zero at both sides.

To complete the proof we have to demonstrate that

$(\underline{PN}\text{-}\underline{MQ})\underline{D}^{-1}(\underline{PN}\text{-}\underline{MQ})^T = \underline{NN}^T + \underline{QVQ}^T \text{-} \underline{NM}^T\underline{P}^{-T}\underline{Q}^T \text{-} \underline{QP}^{-1}\underline{MN}^T$

with $\underline{V} = \underline{P}^{-1}[\underline{MM}^T + (\underline{PN}\text{-}\underline{MQ})\underline{D}^{-1}(\underline{PN}\text{-}\underline{MQ})^T]\underline{P}^{-T}$.

As $\underline{PN}\text{-}\underline{MQ} = \underline{NP}\text{-}\underline{QM}$ we have for the left hand side

$(\underline{NP}\text{-}\underline{QM})\underline{D}^{-1}(\underline{NP}\text{-}\underline{QM})^T = (\underline{N}\text{-}\underline{QMP}^{-1})\underline{PD}^{-1}\underline{P}^T(\underline{N}\text{-}\underline{QMP}^{-1})^T$.

For $\underline{QVQ}^T$ at the right hand side we write:

$$\begin{aligned} \underline{QVQ}^T &= \underline{QP}^{-1}[\underline{MM}^T + (\underline{PN}\text{-}\underline{MQ})\underline{D}^{-1}(\underline{PN}\text{-}\underline{MQ})^T]\underline{P}^{-T}\underline{Q}^T \\ &= \underline{QP}^{-1}\underline{MM}^T\underline{P}^{-T}\underline{Q}^T + \underline{QP}^{-1}(\underline{PN}\text{-}\underline{MQ})\underline{D}^{-1}(\underline{PN}\text{-}\underline{MQ})^T\underline{P}^{-T}\underline{Q}^T \\ &= \underline{QP}^{-1}\underline{MM}^T\underline{P}^{-T}\underline{Q}^T + (\underline{N}\text{-}\underline{QP}^{-1}\underline{M})\underline{QD}^{-1}\underline{Q}^T(\underline{N}\text{-}\underline{QP}^{-1}\underline{M})^T \end{aligned}$$

as $\underline{QP}^{-1}(\underline{PN}\text{-}\underline{MQ}) = \underline{QN}\text{-}\underline{QP}^{-1}\underline{MQ} = \underline{NQ}\text{-}\underline{QMP}^{-1}\underline{Q} = (\underline{N}\text{-}\underline{QMP}^{-1})\underline{Q}$.

The complete right hand side is

$\underline{QP}^{-1}\underline{MM}^T\underline{P}^{-T}\underline{Q}^T + (\underline{N}\text{-}\underline{QP}^{-1}\underline{M})\underline{QD}^{-1}\underline{Q}^T(\underline{N}\text{-}\underline{QP}^{-1}\underline{M})^T + \underline{NN}^T \text{-} \underline{NM}^T\underline{P}^{-T}\underline{Q}^T \text{-} \underline{QP}^{-1}\underline{MN}^T$

or

$(\underline{N}-\underline{Q}\underline{P}^{-1}\underline{M})(I+\underline{Q}\underline{D}^{-1}\underline{Q}^T)(\underline{N}-\underline{Q}\underline{P}^{-1}\underline{M})^T$.

The proof is complete if $\underline{P}\underline{D}^{-1}\underline{P}^T$ is equal to $I+\underline{Q}\underline{D}^{-1}\underline{Q}^T$, which is shown already in lemma 2.1.

To conclude we show that p need not to be equal to q. If $p>q$, the first (p-q) columns of N consists of only zeros and omitting these does not change V as defined in (2.3). If $p<q$, Q has (q-p) zero columns, but these do not influence $\bar{Q}\bar{Q}^T$. □

Corollary 1

The covariance matrix (apart of σ^2) for the MA(q) model is

$$V_{MA} = NN^T + MM^T. \tag{2.5}$$

Corollary 2

The covariance matrix (apart of σ^2) for the AR(p) model is

$$V_{AR} = (P^TP - QQ^T)^{-1}. \tag{2.6}$$

Proof

The proof of Corollary 1 is trivial. Substituting $P=I$ and $Q=0$ in (2.2), we get the MA(q) expression for V. To prove Corollary 2 use (2.3); for $q=0$ we have $[N\ \ M]=I_T$ and $\bar{P}=B_{T+q,T+q}(1,\vartheta_1,..,\vartheta_p,0,..0)=P$ and $\bar{Q}=B^T_{p,T+q}(\vartheta_p,..,\vartheta_1)=Q$.
□

It is clear that the second term of (2.4) within brackets is of rank $r=\max(p,q)$. Because of the commuting property, PN-MQ can be written as $\begin{bmatrix}\underline{PN\text{-}MQ}\\ 0\end{bmatrix}$, where the matrices in the upper part are square of order $r\times r$. This makes clear that the main part of V consists of $P^{-1}MM^TP^{-T}$, the rest being a correction matrix of which the order is max(p,q). Furthermore (2.4) is easy to invert: the core of the inverse consists of a $(r\times r)$ matrix, which can be triangulized. Using an expression for the inverse of the sum of two matrices (see *e.g.* Rao, 1973, p. 33), we get

$$V^{-1} = P^TM^{-T}\{I_T - R(R^TR+\underline{D})^{-1}R^T\}M^{-1}P \tag{2.7}$$

with $R = M^{-1}PN-Q \in \mathbb{R}^{T\times r}$. It is not clear whether it is possible to write (2.7) in a form similar to (2.6), where the MA part and the AR part are separated.

The determinant of V can be obtained in the following way. Observing that the value of the determinant of $M^{-1}P$ is equal to one, we have

$$|V| = |I_T + M^{-1}\begin{bmatrix}\underline{PN}\text{-}\underline{MQ}\\ 0\end{bmatrix}\underline{D}^{-1}\begin{bmatrix}\underline{PN}\text{-}\underline{MQ}\\ 0\end{bmatrix}^T M^{-T}|$$

or

$$|V| = |I_r + \underline{D}^{-1}(\underline{PN}\text{-}\underline{MQ})^T M^{*T} M^{*}(\underline{PN}\text{-}\underline{MQ})| \tag{2.8}$$

where $M^* \in \mathbb{R}^{T\times r}$ consists of the first r columns of M^{-1}. The equality is due to the fact, that the second term of the sum in both equations has the same nonzero characteristic roots. The evaluation of the determinant can thus be reduced from a (T×T) matrix to one of dimensions (r×r), the higher number of AR or MA parameters. Moreover it is easy to see, that in the AR case the determinant reduces to $|\underline{D}^{-1}|$, which is independent of T. Then $\underline{M} = I_p$ and $\underline{N}$ disappears and the determinant becomes
$|I_p + (\underline{P}^T\underline{P}\text{-}\underline{QQ}^T)^{-1}\underline{Q}^T\underline{Q})| = |(\underline{P}^T\underline{P}\text{-}\underline{QQ}^T)^{-1}||\underline{P}^T\underline{P}\text{-}\underline{QQ}^T + \underline{Q}^T\underline{Q})| = |(\underline{P}^T\underline{P}\text{-}\underline{QQ}^T)^{-1}|$
as $|\underline{P}^T\underline{P}\text{-}\underline{QQ}^T + \underline{Q}^T\underline{Q}| = |\underline{PP}^T| = I_p$.

2.5 Some examples

Next we will show some examples of a few ARMA covariance matrices. While the inverted part is simple, the covariance matrix itself is certainly not. Generally, the complexity of the elements of the covariance matrix grows as their distance from the main diagonal is larger and the number of AR parameters is larger. We present at most only the first eight elements of the first row (or column) of the covariance matrix unless these elements are zero. The results are obtained using *Mathematics* and some further editing.

For the most simple case, MA(1) we get only three diagonals. The main diagonal has as elements $1+\alpha_1^2$, the side diagonals α_1. For MA(2) we have one main diagonal with elements $1+\alpha_1^2+\alpha_2^2$ and two side diagonals with elements $\alpha_1+\alpha_1\alpha_2$ for the first one and α_2 for the second one. The result for the AR(1) case is well-known: 1, $-\vartheta_1$, ϑ_1^2, $-\vartheta_1^3$, ϑ_1^4, $-\vartheta_1^5$, and so forth, where every element has to be divided by the determinant of $P^TP\text{-}QQ^T$, *viz.* $1\text{-}\vartheta_1^2$.

For the AR(2) the determinant is $1 - \vartheta_1^2 + 2\,\vartheta_1^2\,\vartheta_2 - 2\,\vartheta_2^2 - \vartheta_1^2\,\vartheta_2^2 + \vartheta_2^4$ and the elements of the first row become:

$1 - \vartheta_2^2$

$-(1 - \vartheta_2)\ \vartheta_1$

$(1 - \vartheta_2)\ (\vartheta_1^2 - \vartheta_2 - \vartheta_2^2)$

$-(1 - \vartheta_2)\ (\vartheta_1^3 - 2\ \vartheta_1\ \vartheta_2 - \vartheta_1\ \vartheta_2^2)$

$(1 - \vartheta_2)\ (\vartheta_1^4 - 3\ \vartheta_1^2\ \vartheta_2 - \vartheta_1^2\ \vartheta_2^2 + \vartheta_2^2 + \vartheta_2^3)$

$-(1 - \vartheta_2)\ (\vartheta_1^5 - 4\ \vartheta_1^3\ \vartheta_2 - \vartheta_1^3\ \vartheta_2^2 + 3\ \vartheta_1\ \vartheta_2^2 + 2\ \vartheta_1\ \vartheta_2^3)$

$(1 - \vartheta_2)\ (\vartheta_1^6 - 5\ \vartheta_1^4\ \vartheta_2 - \vartheta_1^4\ \vartheta_2^2 + 6\ \vartheta_1^2\ \vartheta_2^2 + 3\ \vartheta_1^2\ \vartheta_2^3 - \vartheta_2^3 - \vartheta_2^4)$

$-(1 - \vartheta_2)\ (\vartheta_1^7 - 6\ \vartheta_1^5\ \vartheta_2 - \vartheta_1^5\ \vartheta_2^2 + 10\ \vartheta_1^3\ \vartheta_2^2 + 4\ \vartheta_1^3\ \vartheta_2^3 - 4\ \vartheta_1\ \vartheta_2^3 - 3\ \vartheta_1\ \vartheta_2^4)$

For the ARMA(1,1) we get as first row, to be divided by $1 - \vartheta_1^2$:

$1 - 2\ \vartheta_1\ \alpha_1 + \alpha_1^2$

$-\ (\vartheta_1 - \alpha_1)\ (1 - \vartheta_1\ \alpha_1)$

$\vartheta_1\ (\vartheta_1 - \alpha_1)\ (1 - \vartheta_1\ \alpha_1)$

$-\vartheta_1^2\ (\vartheta_1 - \alpha_1)\ (1 - \vartheta_1\ \alpha_1)$

$\vartheta_1^3\ (\vartheta_1 - \alpha_1)\ (1 - \vartheta_1\ \alpha_1)$

$-\vartheta_1^4\ (\vartheta_1 - \alpha_1)\ (1 - \vartheta_1\ \alpha_1)$

$\vartheta_1^5\ (\vartheta_1 - \alpha_1)\ (1 - \vartheta_1\ \alpha_1)$

$-\vartheta_1^6\ (\vartheta_1 - \alpha_1)\ (1 - \vartheta_1\ \alpha_1).$

For AR(1,2) we have again as determinant of the AR part $1 - \vartheta_1^2$, and for the elements of the covariance matrix we first define $A=(\vartheta_1^2 - \vartheta_1\ \alpha_1 + \alpha_2)$, $B=(1 - \vartheta_1\ \alpha_1 + \vartheta_1^2\ \alpha_2)$, $C=(\vartheta_1 - \alpha_1 + \vartheta_1\alpha_2)$. For the first eight elements the result is:

$\alpha_2 A + B - \alpha_1 C$

$-\alpha_2\ C + (\alpha_1 - \vartheta_1)\ B$

$A\ B$

$-\vartheta_1\ A\ B$

$\vartheta_1^2\ A\ B$

$-\vartheta_1^3\ A\ B$

$\vartheta_1^4\ A\ B$

$-\vartheta_1^5\ A\ B$

For the ARMA(2,1) covariance matrix we have as divider $1 - \vartheta_1^2 + 2\ \vartheta_1^2\ \vartheta_2 - 2\ \vartheta_2^2 - \vartheta_1^2\ \vartheta_2^2 + \vartheta_2^4$. We use $A=(-1+\vartheta_2)$, $B=1+\alpha^2$, $C=\vartheta_1\alpha$ and $D=\alpha(1-\vartheta_2^2)$ to present the first eight elements of the covariance matrix:

$$-A(-B\ \vartheta_1 + C\ \vartheta_1 + \alpha D\ )$$

$$A(-B(\ \vartheta_1^2 - \vartheta_2^2 - \vartheta_2\ ) + C(\ \vartheta_1^2 - \vartheta_2^2 - 2\ \vartheta_2 + 1)\ )$$

$$-A(-B(\ \vartheta_1^3 - \vartheta_1\vartheta_2^2 - 2\ \vartheta_1\ \vartheta_2) + C(\ \vartheta_1^3 - \vartheta_1\vartheta_2^2 - 3\ \vartheta_1\ \vartheta_2 + \vartheta_1) - \vartheta_2 D\ )$$

$$A\ (-B(\ \vartheta_1^4 - \vartheta_1^2\ \vartheta_2^2 - 3\ \vartheta_1^2\ \vartheta_2 + \vartheta_2^3 + \vartheta_2^2\ ) + \\ + C(\ \vartheta_1^4 - \vartheta_1^2\ \vartheta_2^2 - 4\ \vartheta_1^2\ \vartheta_2 + \vartheta_1^2 + 2\ \vartheta_2^3 + 2\ \vartheta_2^2 - 2\ \vartheta_2\ ))$$

$$-A(-B(\ \vartheta_1^5 - \vartheta_1^3\ \vartheta_2^2 - 4\ \vartheta_1^3\ \vartheta_2 + 2\ \vartheta_1\ \vartheta_2^3 + 3\ \vartheta_1\ \vartheta_2^2\ ) + \\ + C(\ \vartheta_1^5 - \vartheta_1^3\ \vartheta_2^2 - 5\ \vartheta_1^3\ \vartheta_2 + \vartheta_1^3 + 3\ \vartheta_1\ \vartheta_2^3 + 5\ \vartheta_1\ \vartheta_2^2 - 3\ \vartheta_1\ \vartheta_2\ ) + \vartheta_2^2 D)$$

$$A(-B(\ \vartheta_1^6 - \vartheta_1^4\ \vartheta_2^2 - 5\ \vartheta_1^4\ \vartheta_2 + 3\ \vartheta_1^2\ \vartheta_2^3 + 6\ \vartheta_1^2\ \vartheta_2^2 - \vartheta_2^4 - \vartheta_2^3\ ) + \\ + C(\ \vartheta_1^6 - \vartheta_1^4\ \vartheta_2^2 - 6\ \vartheta_1^4\ \vartheta_2 + \vartheta_1^4 + 4\ \vartheta_1^2\ \vartheta_2^3 + 9\ \vartheta_1^2\ \vartheta_2^2 - 4\ \vartheta_1^2\ \vartheta_2 - 3\ \vartheta_2^4 - 2\ \vartheta_2^3 + 3\ \vartheta_2^2\ ))$$

$$-A(-B(\ \vartheta_1^7 - \vartheta_1^5\ \vartheta_2^2 - 6\ \vartheta_1^5\ \vartheta_2 + 4\ \vartheta_1^3\ \vartheta_2^3 + 10\ \vartheta_1^3\ \vartheta_2^2 - 3\ \vartheta_1\ \vartheta_2^4 - 4\ \vartheta_1\ \vartheta_2^3\ ) \\ + C(\ \vartheta_1^7 - \vartheta_1^5\ \vartheta_2^2 - 7\ \vartheta_1^5\ \vartheta_2 + \vartheta_1^5 + 5\ \vartheta_1^3\ \vartheta_2^3 + 14\ \vartheta_1^3\ \vartheta_2^2 - 5\ \vartheta_1^3\ \vartheta_2 - 6\ \vartheta_1\ \vartheta_2^4 - 7\ \vartheta_1\ \vartheta_2^3 + 6\ \vartheta_1\ \vartheta_2^2\ ) - \vartheta_2^3 D\ )$$

The ARMA(2,2) has as determinant for the AR part $1 - \vartheta_1^2 + 2\ \vartheta_1^2\ \vartheta_2 - 2\ \vartheta_2^2 - \vartheta_1^2\ \vartheta_2^2 + \vartheta_2^4$.
To show the growing complexity we give the first elements of its covariance matrix, without any attempt to beautify:

$$(-1 + \vartheta_2)\ (-1 - \vartheta_2 + 2\ \vartheta_1\ \alpha_1 - \alpha_1^2 - \vartheta_2\ \alpha_1^2 - 2\ \vartheta_1^2\ \alpha_2 + 2\ \vartheta_2\ \alpha_2 + 2\ \vartheta_2^2\ \alpha_2 + 2\ \vartheta_1\ \alpha_1\ \alpha_2 - \alpha_2^2 - \vartheta_2\ \alpha_2^2\ )$$

$$(-1 + \vartheta_2)\ (\vartheta_1 - \alpha_1 - \vartheta_1^2\ \alpha_1 + \vartheta_2^2\ \alpha_1 + \vartheta_1\ \alpha_1^2 + \vartheta_1\ \alpha_2 + \vartheta_1^3\ \alpha_2 - 2\ \vartheta_1\ \vartheta_2\ \alpha_2 - \vartheta_1\ \vartheta_2^2\ \alpha_2 - \alpha_1\ \alpha_2 - \vartheta_1^2\ \alpha_1\ \alpha_2 + \vartheta_2^2$$
$$\alpha_1\ \alpha_2 + \vartheta_1\ \alpha_2^2\ )$$

$$(-1 + \vartheta_2)\ (-\vartheta_1^2 + \vartheta_2 + \vartheta_2^2\ ) + \vartheta_1\ (-1 + \vartheta_2)\ (\vartheta_1^2 - 2\ \vartheta_2 - \vartheta_2^2\ )\ \alpha_1 - (-1 + \vartheta_2)\ (\vartheta_1^4 - 3\ \vartheta_1^2\ \vartheta_2 + \vartheta_2^2 - \vartheta_1^2\ \vartheta_2^2 + \vartheta_2^3)$$
$$\alpha_2 + \alpha_1\ (\vartheta_1\ (-1 + \vartheta_2) + (-1 + \vartheta_2)\ (-\vartheta_1^2 + \vartheta_2 + \vartheta_2^2\ )\ \alpha_1 + \vartheta_1\ (-1 + \vartheta_2)\ (\vartheta_1^2 - 2\ \vartheta_2 - \vartheta_2^2\ )\ \alpha_2) +$$
$$\alpha_2\ (1 - \vartheta_2^2 + \vartheta_1\ (-1 + \vartheta_2)\ \alpha_1 + (-1 + \vartheta_2)\ (-\vartheta_1^2 + \vartheta_2 + \vartheta_2^2\ )\ \alpha_2)$$

$$(-1 + \vartheta_2)\ (\vartheta_1^3 - 2\ \vartheta_1\ \vartheta_2 - \vartheta_1\ \vartheta_2^2 - \vartheta_1^2\ \alpha_1 - \vartheta_1^4\ \alpha_1 + \vartheta_2\ \alpha_1 + 3\ \vartheta_1^2\ \vartheta_2\ \alpha_1 + \vartheta_1^2\ \vartheta_2^2\ \alpha_1 - \vartheta_2^3\ \alpha_1 + \vartheta_1^3\ \alpha_1^2 - 2\ \vartheta_1\ \vartheta_2$$
$$\alpha_1^2 - \vartheta_1\ \vartheta_2^2\ \alpha_1^2 + \vartheta_1\ \alpha_2 + \vartheta_1^5\ \alpha_2 - 4\ \vartheta_1^3\ \vartheta_2\ \alpha_2 + 3\ \vartheta_1\ \vartheta_2^2\ \alpha_2 - \vartheta_1^3\ \vartheta_2^2\ \alpha_2 + 2\ \vartheta_1\ \vartheta_2^3\ \alpha_2 - \vartheta_1^2\ \alpha_1\ \alpha_2 - \vartheta_1^4\ \alpha_1\ \alpha_2 + \vartheta_2$$
$$\alpha_1\ \alpha_2 + 3\ \vartheta_1^2\ \vartheta_2\ \alpha_1\ \alpha_2 + \vartheta_1^2\ \vartheta_2^2\ \alpha_1\ \alpha_2 - \vartheta_2^3\ \alpha_1\ \alpha_2 + \vartheta_1^3\ \alpha_2^2 - 2\ \vartheta_1\ \vartheta_2\ \alpha_2^2 - \vartheta_1\ \vartheta_2^2\ \alpha_2^2\ )$$

$$-((-1 + \vartheta_2)\ (\vartheta_1^4 - 3\ \vartheta_1^2\ \vartheta_2 + \vartheta_2^2 - \vartheta_1^2\ \vartheta_2^2 + \vartheta_2^3\ )) + \vartheta_1\ (-1 + \vartheta_2)\ (\vartheta_1^4 - 4\ \vartheta_1^2\ \vartheta_2 + 3\ \vartheta_2^2 - \vartheta_1^2\ \vartheta_2^2 + 2\ \vartheta_2^3\ )\ \alpha_1 -$$
$$(-1 + \vartheta_2)\ (\vartheta_1^6 - 5\ \vartheta_1^4\ \vartheta_2 + 6\ \vartheta_1^2\ \vartheta_2^2 - \vartheta_1^4\ \vartheta_2^2 - \vartheta_2 3 + 3\ \vartheta_1^2\ \vartheta_2^3 - \vartheta_2^4\ )\ \alpha_2 + \alpha_2\ ((-1 + \vartheta_2)\ (-\vartheta_1^2 + \vartheta_2 + \vartheta_2^2\ ) + \vartheta_1$$
$$(-1 + \vartheta_2)\ (\vartheta_1^2 - 2\ \vartheta_2 - \vartheta_2^2\ )\ \alpha_1 - (-1 + \vartheta_2)\ (\vartheta_1^4 - 3\ \vartheta_1^2\ \vartheta_2 + \vartheta_2^2 - \vartheta_1^2\ \vartheta_2^2 + \vartheta_2^3\ )\ \alpha_2) + \alpha_1\ (\vartheta_1\ (-1 + \vartheta_2)$$
$$(\vartheta_1^2 - 2\ \vartheta_2 - \vartheta_2^2\ ) - (-1 + \vartheta_2)\ (\vartheta_1^4 - 3\ \vartheta_1^2\ \vartheta_2 + \vartheta_2^2 - \vartheta_1^2\ \vartheta_2^2 + \vartheta_2^3\ )\ \alpha_1 + \vartheta_1\ (-1 + \vartheta_2)$$
$$(\vartheta_1^4 - 4\ \vartheta_1^2\ \vartheta_2 + 3\ \vartheta_2^2 - \vartheta_1^2\ \vartheta_2^2 + 2\ \vartheta_2^3\ )\ \alpha_2)$$

$$(-1 + \vartheta_2)\ (\vartheta_1^5 - 4\ \vartheta_1^3\ \vartheta_2 + 3\ \vartheta_1\ \vartheta_2^2 - \vartheta_1^3\ \vartheta_2^2 + 2\ \vartheta_1\ \vartheta_2^3 - \vartheta_1^4\ \alpha_1 - \vartheta_1^6\ \alpha_1 + 3\ \vartheta_1^2\ \vartheta_2\ \alpha_1 + 5\ \vartheta_1^4\ \vartheta_2\ \alpha_1 - \vartheta_2^2\ \alpha_1 -$$
$$5\ \vartheta_1^2\ \vartheta_2^2\ \alpha_1 + \vartheta_1^4\ \vartheta_2^2\ \alpha_1 - 3\ \vartheta_1^2\ \vartheta_2^3\ \alpha_1 + \vartheta_2^4\ \alpha_1 + \vartheta_1^5\ \alpha_1^2 - 4\ \vartheta_1^3\ \vartheta_2\ \alpha_1^2 + 3\ \vartheta_1\ \vartheta_2^2\ \alpha_1^2 - \vartheta_1^3\ \vartheta_2^2\ \alpha_1^2 + 2\ \vartheta_1\ \vartheta_2^3\ \alpha_1^2 +$$
$$\vartheta_1^3\ \alpha_2 + \vartheta_1^7\ \alpha_2 - 2\ \vartheta_1\ \vartheta_2\ \alpha_2 - 6\ \vartheta_1^5\ \vartheta_2\ \alpha_2 - \vartheta_1\ \vartheta_2^2\ \alpha_2 + 10\ \vartheta_1^3\ \vartheta_2^2\ \alpha_2 - \vartheta_1^5\ \vartheta_2^2\ \alpha_2 - 4\ \vartheta_1\ \vartheta_2^3\ \alpha_2 + 4\ \vartheta_1^3\ \vartheta_2^3\ \alpha_2 -$$
$$3\ \vartheta_1\ \vartheta_2^4\ \alpha_2 - \vartheta_1^4\ \alpha_1\ \alpha_2 - \vartheta_1^6\ \alpha_1\ \alpha_2 + 3\ \vartheta_1^2\ \vartheta_2\ \alpha_1\ \alpha_2 + 5\ \vartheta_1^4\ \vartheta_2\ \alpha_1\ \alpha_2 - \vartheta_2^2\ \alpha_1\ \alpha_2 - 5\ \vartheta_1^2\ \vartheta_2^2\ \alpha_1\ \alpha_2 + \vartheta_1^4\ \vartheta_2^2\ \alpha_1\ \alpha_2$$
$$-\ 3\ \vartheta_1^2\ \vartheta_2^3\ \alpha_1\ \alpha_2 + \vartheta_2^4\ \alpha_1\ \alpha_2 + \vartheta_1^5\ \alpha_2^2 - 4\ \vartheta_1^3\ \vartheta_2\ \alpha_2^2 + 3\ \vartheta_1\ \vartheta_2^2\ \alpha_2^2 - \vartheta_1^3\ \vartheta_2^2\ \alpha_2^2 + 2\ \vartheta_1\ \vartheta_2^3\ \alpha_2^2\ ).$$

2.6 Conclusion

In this chapter we used a matrix approach for the ARMA error structure. This technique appears fruitful. It enables us to derive a closed form expression for the ARMA covariance matrix. The main result is found Theorem 2.3. The covariance matrix can be written as a function of four matrices, each with a simple triangular structure. Further results are expressions for the inverse and the determinant of the covariance matrix. Especially the determinant of a pure AR covariance matrix is simple: it can be computed from the $p \times p$ upper left sub matrix and it is independent of the order of the covariance matrix. The relationship between the stationary condition - usually expressed in polynomial form - and the matrix form we propose, can be found in Theorem 2.1.

The result was found in a two step procedure: first a matrix equation with the covariance matrix as unknown was derived, secondly this equation was solved. The solution is trivial for the MA case and rather easy for the AR case. The main problem was to find the general solution.
The simple triangular structure of the building blocks will enable us to derive analytical expressions for the first and second derivatives. This in turn will be used to investigate the ML function.

III LAG FORMS AND DIFFERENTIALS

3.1 Lag matrix

The form of the covariance matrix as presented in the previous chapter is not suitable for differentiation because it is not written explicitely as a function of the parameters of interest, α and ϑ. Therefore we will develop an alternative form where this is indeed the case. The 'building' blocks (P, Q, M, and N) are rather simple matrices, which after differentiating, consist of one diagonal of ones while the rest is zero. This property will be used to write the covariance matrix or at least parts of it explicitely as function of the parameters. To do so we first define a so called lag matrix, which permits an alternative expression for the covariance matrix. Moreover we give some of its properties.

3.1.1 Definition

Let ι_h be a $T \times 1$ vector of which all elements are zero, apart from element h, which is 1. Define the $T \times T$ matrix $L_k(n,m) = \sum_{h=n+1}^{T-m+k} \iota_h \iota_{h-k}^T$, with $|k| \leq T-1$, $\max(0,k) \leq n \leq T$, $\max(0,k) \leq m \leq T$. We write $L_k = L_k(0,0)$ and if $n=m$ we will use $L_k(n) = L_k(n,n)$.

This definition of L implies that L has one (off-)diagonal consisting of ones, all other elements being zero. If $n=m=0$, then all elements of this diagonal are one. We allow the first or last elements of this diagonal to be zero. To define which elements are zero we have the choice between the number of first rows and the last columns on the one hand, or first columns and last rows on the other. We take, arbitrarily, the first way. This means that n and m are positive in case k is positive. If $n=m=k=0$ we get the identity matrix. If k is positive L can be regarded as a lag matrix, with the same property as the usual lag operator. Some typical examples for $T=10$ are :

$$L_{-2}(0,3) = \begin{bmatrix} . & . & 1 & & & & & & & \\ & . & . & 1 & & & & & & \\ & & . & . & 1 & & & & & \\ & & & . & . & 1 & & & & \\ & & & & . & . & 1 & & & \\ & & & & & . & . & . & & \\ & & & & & & . & . & . & \\ & & & & & & & . & . & . \\ & & & & & & & & . & . \\ & & & & & & & & & . \end{bmatrix} \quad \text{and } L_3(3,4) = \begin{bmatrix} . & & & & & & & & & \\ . & . & & & & & & & & \\ . & . & . & & & & & & & \\ 1 & . & . & . & & & & & & \\ & 1 & . & . & . & & & & & \\ & & 1 & . & . & . & & & & \\ & & & 1 & . & . & . & & & \\ & & & & 1 & . & . & . & & \\ & & & & & 1 & . & . & . & \\ & & & & & & & . & . & . \end{bmatrix}$$

All elements not being 1 are zero. The dots are only printed to make clear which diagonal is meant and which elements are not ones.

Let $z \in \mathbb{R}^{T\times 1}$. Then, for $k \geq 0$:

$$L_k(k)z = \sum_{h=1+k}^{T} \iota_h \iota_{h-k}^T z = \sum_{h=1+k}^{T} z_{h-k}\iota_h = (\underbrace{0 \ldots 0}_{\leftarrow k \rightarrow} \; \underbrace{z_1 \ldots z_{T-k}}_{\leftarrow T-k \rightarrow})^T.$$

If n or m is greater than k the n-k first elements or m-k last elements disappear. We do, however, not exclude negative values for k. In this sense k is not an ordinary lag matrix. In the next sections we give some properties of L.

3.1.2 Properties

Lemma 3.1

Transposition: $L_k^T(n,m) = L_{-k}(n-k,m-k)$.

Proof

Straightforward transposition of L and and changing the index h to j=h-k we get:

$$L_k^T(n,m) = \Big(\sum_{h=1+n}^{T-m+k} \iota_h \iota_{h-k}^T \Big)^T = \sum_{h=1+n}^{T-m+k} \iota_{h-k} \iota_h^T = \sum_{j=1+n-k}^{T-(m-k)-k} \iota_j \iota_{j+k}^T = L_{-k}(n-k,m-k). \ \square$$

Lemma 3.2

Multiplication:

$$L_{k_1}(n_1,m_1) \cdot L_{k_2}(n_2,m_2) = L_{k_1+k_2}(\max(n_1,n_2+k_1),\max(m_1+k_2,m_2)).$$

Proof

$$L_{k_1}(n_1,m_1)\cdot L_{k_2}(n_2,m_2) =$$

$$= \sum_{h=1+n_1}^{T-m_1+k_1} \iota_h \iota_{h-k_1}^T \sum_{g=1+n_2}^{T-m_2+k_2} \iota_g \iota_{g-k_2}^T$$

$$= \sum_{h=1+n_1}^{T-m_1+k_1} \sum_{g=1+n_2}^{T-m_2+k_2} \iota_h \iota_{h-k_1}^T \iota_g \iota_{g-k_2}^T$$

$$= \sum_{j=1+n_1}^{T-m_1+k_1} \sum_{j=1+n_2+k_1}^{T-m_2+k_2+k_1} \iota_j \iota_{j-k_1}^T \iota_{j-k_1} \iota_{j-k_1-k_2}^T.$$

The expression $\iota_{h-k_1}^T \iota_g$ is only non-zero if $g=h-k_1$, or $h=g+k_1=j$. Here j runs from max $(1+n_1, 1+n_2+k_1)$ to min $(T-m_1+k_1, T-m_2+k_2+k_1)$, which is the same as from $1+\max(n_1,n_2+k_1)$ to $T-\max(m_1+k_2,m_2)+k_1+k_2$, while the lag is equal to k_1+k_2. □

3.2 Expressions in lag form

Next we present several expressions for the covariance matrix or parts of it in lag form. From this form we easily derive the corresponding differential which will be used in the sequel to find the derivatives. First we will treat the differentials for the MA parameters, next for the AR parameters.

For the MA part of the covariance matrix or the MA covariance matrix itself we need an easy to differentiate expression for [N M]. Denote zero-one vectors of length $T+q$ with a bar.

Proposition 3.1

$$[N\ M] = \sum_{i=0}^{q} \sum_{k=1}^{T} \iota_k \bar{\iota}_{k+q-i}^T \alpha_i.$$

Proof

From $\bar{M} = B_{T+q,T+q}(1,\alpha_1,..,\alpha_q,0,..0)$ we see that $[N\ M] = [0_{T,q}\ I_T]\bar{M}$ and $\bar{M} = \sum_{i=0}^{q} L_i(i)\alpha_i = \sum_{i=0}^{q}\sum_{h=1+i}^{T+q} \bar{\iota}_h\bar{\iota}_{h-i}^T\alpha_i$, while $[0_{T,q}\ I_T] = \sum_{h=1}^{T}\iota_h\bar{\iota}_{h+q}^T$. Then

$$[N\ M] = \Big(\sum_{g=1}^{T}\iota_g\bar{\iota}_{g+q}^T\Big)\Big(\sum_{i=0}^{q}\sum_{h=1}^{T+q}\bar{\iota}_h\bar{\iota}_{h-i}^T\alpha_i\Big)$$

$$= \sum_{i=0}^{q}\sum_{g=1}^{T}\sum_{h=1}^{T+q}\iota_g\bar{\iota}_{g+q}^T\bar{\iota}_h\bar{\iota}_{h-i}^T\alpha_i$$

$$= \sum_{i=0}^{q}\sum_{k=1}^{T}\iota_k\bar{\iota}_{k+q-i}^T\alpha_i$$

as $\bar{\iota}_{g+q}^T\bar{\iota}_h = 0$, unless $k = g = h-q$, which implies $h-i = k+q-i$, while k runs from $\max(1,1-q) = 1$ to T. □

For the MA covariance matrix V_{MA} we get:

Proposition 3.2

$$V_{MA} = \sum_{i=0}^{q}\Big(\sum_{j=0}^{i} L_{j-i}^T\alpha_j + \sum_{j=i+1}^{q} L_{i-j}\alpha_j\Big)\alpha_i.$$

Proof

From corrolary 2.2 and proposition 3.1 we conclude

$$V_{MA} = [N\ M][N\ M]^T$$

$$= \Big(\sum_{i=0}^{q}\sum_{h=1}^{T}\iota_h\bar{\iota}_{h+q-i}^T\alpha_i\Big)\Big(\sum_{j=0}^{q}\sum_{k=1}^{T}\iota_k\bar{\iota}_{k+q-j}^T\alpha_j\Big)^T$$

$$= \sum_{i=0}^{q}\sum_{j=0}^{q}\sum_{h=1}^{T}\sum_{k=1}^{T}\iota_h\bar{\iota}_{h+q-i}^T\iota_{k+q-j}\bar{\iota}_k^T\alpha_i\alpha_j$$

Observe that $\bar{\iota}_{h+q-i}^T\bar{\iota}_{k+q-j} = 0$, unless $h-i = k-j$. Use this to change from the indices h and k to g. Let $g = h$, then we have for the remaining indices $h+q-i = k+q-j = g+q-i$, $k = g-i+j$. As $g = h$ and $g = k+i-j$, g runs from $\max(1,1+i-j)$ to $\min(T,T+i-j)$, or from $1+\max(i-j,0)$ to $T+i-j-\max(i-j,0)$. Hence

$$V_{MA} = \sum_{i=0}^{q}\sum_{j=0}^{q}\sum_{g=1+\max(0,i-j)}^{T+i-j-\max(0,i-j)}\iota_g\iota_{g-i+j}^T\alpha_i\alpha_j$$

$$= \sum_{i=0}^{q} \sum_{j=0}^{q} L_{i-j}(\max(i-j,0)\alpha_i\alpha_j$$

$$= \sum_{i=0}^{q} \quad \sum_{j=0}^{i} L^T_{j-i}\alpha_j + \sum_{j=i+1}^{q} L_{i-j}\alpha_j)\alpha_i. \ \square$$

Proposition 3.3

Let $\bar{V}_{AR} = (\bar{P}^T\bar{P}\text{-}\bar{Q}\bar{Q}^T)^{-1}$ and $\bar{V}_{AR}[i]$ the determining element of its i^{th} diagonal. The differential of V for α is

$$\mathbf{d}V = \sum_{i=1}^{q} \sum_{j=0}^{q} \sum_{h=1}^{T} \sum_{k=1}^{T} \bar{V}_{AR}[h\text{-}k\text{-}i+j](\iota_h\iota_k^T + \iota_k\iota_h^T)\alpha_j\mathbf{d}\alpha_i.$$

Proof

From (2.3) and proposition 3.1 we get

$$\mathbf{d}V = \mathbf{d}([N\ \ M]\bar{V}_{AR}[N\ \ M]^T)$$

$$= \mathbf{d}([N\ \ M])\bar{V}_{AR}[N\ \ M]^T + [N\ \ M]\bar{V}_{AR}\mathbf{d}[N\ \ M]^T))$$

$$= (\sum_{i=0}^{q} \sum_{h=1}^{T} \iota_h\bar{\iota}^T_{h+q-i}\mathbf{d}\alpha_i)\bar{V}_{AR}(\sum_{j=0}^{q} \sum_{k=1}^{T} \iota_k\bar{\iota}^T_{k+q-j}\alpha_j)^T +$$

$$(\sum_{i=0}^{q} \sum_{h=1}^{T} \iota_h\bar{\iota}^T_{h+q-i}\alpha_i)\bar{V}_{AR}(\sum_{j=0}^{q} \sum_{k=1}^{T} \iota_k\bar{\iota}^T_{k+q-j}\mathbf{d}\alpha_j)^T$$

$$= \sum_{i=0}^{q} \sum_{j=0}^{q} \sum_{h=1}^{T} \sum_{k=1}^{T} \iota_h\bar{\iota}^T_{h+q-i}\bar{V}_{AR}\bar{\iota}_{k+q-j}\iota^T_k\alpha_j\mathbf{d}\alpha_i +$$

$$\sum_{j=0}^{q} \sum_{i=0}^{q} \sum_{k=1}^{T} \sum_{h=1}^{T} \iota_k\bar{\iota}^T_{k+q-j}\bar{V}_{AR}\bar{\iota}_{h+q-i}\iota^T_h\alpha_j\mathbf{d}\alpha_i$$

$$= \sum_{i=0}^{q} \sum_{j=0}^{q} \sum_{h=1}^{T} \sum_{k=1}^{T} (\iota_h\bar{V}_{AR}[h\text{-}k\text{-}i+j]\iota_k^T + \iota_k\bar{V}_{AR}[k\text{-}h\text{-}j+i]\iota_h^T)\alpha_j\mathbf{d}\alpha_i$$

$$= \sum_{i=1}^{q} \sum_{j=0}^{q} \sum_{h=1}^{T} \sum_{k=1}^{T} \bar{V}_{AR}[h\text{-}k\text{-}i+j](\iota_h\iota_k^T + \iota_k\iota_h^T)\alpha_j\mathbf{d}\alpha_i, \text{ as } \mathbf{d}\alpha_0 = 0. \ \square$$

There seems no easy way to express the AR covariance matrix in lag form, but for its inverse we have the following result.

Proposition 3.4

If $T \geq 2p$, $V_{AR}^{-1} = \sum_{i=0}^{p} \sum_{j=0}^{p} L_{j-i}(j)\vartheta_i\vartheta_j$.

Proof

Use lag forms to write the square matrix $P = \sum_{i=0}^{p} L_i(i)\vartheta_i$. To express the $(T{\times}p)$ matrix Q in terms of lag matrices, define the $(T{\times}T)$-matrix $\tilde{Q} = \sum_{i=0}^{p} L_{i-p}\vartheta_i$ and $Q = \tilde{Q}[I_p \;\; O_{p,T-p}]^T$. Substituting these forms in (2.6) we get

$$V_{AR}^{-1} = (\sum_{i=0}^{p} L_i(i)\vartheta_i)^T(\sum_{j=0}^{p} L_j(j)\vartheta_j) - (\sum_{i=0}^{p} L_{i-p}\vartheta_i)[I_p \;\; O_{p,T-p}]^T[I_p \;\; O_{p,T-p}](\sum_{j=0}^{p} L_{j-p}\vartheta_j$$

The part before the minus sign is

$$(\sum_{i=0}^{p} L_i(i)\vartheta_i)^T(\sum_{j=0}^{p} L_j(j)\vartheta_j) =$$

$$\sum_{i=0}^{p} \sum_{j=0}^{p} L_{-i}L_j(j)\vartheta_i\vartheta_j = \sum_{i=0}^{p} \sum_{j=0}^{p} L_{j-i}(\max(0,i-j),j)\vartheta_i\vartheta_j.$$

For the second part first observe, that $[I_p \;\; O_{p,T-p}]^T[I_p \;\; O_{p,T-p}] = L_0(0,T-p)$ and we get

$$(\sum_{i=0}^{p} L_{i-p}\vartheta_i)L_0(0,T-p)(\sum_{j=0}^{p} L_{j-p}\vartheta_j)^T =$$

$$= \sum_{i=0}^{p} \sum_{j=0}^{p} L_{i-p}L_0(0,T-p)L_{p-j}(p-j)\vartheta_i\vartheta_j$$

$$= \sum_{i=0}^{p} \sum_{j=0}^{p} L_{i-p}(0,T-p)L_{p-j}(p-j)\vartheta_i\vartheta_j$$

$$= \sum_{i=0}^{p} \sum_{j=0}^{p} L_{i-j}(\max(0,-j+i),T-j)\vartheta_i\vartheta_j$$

$$= \sum_{i=0}^{p} \sum_{j=0}^{p} L_{j-i}(\max(0,j-i),T-i)\vartheta_j\vartheta_i.$$

The lag matrices of the first and second part are $L_{j-i}(\max(0,i-j),j)$ and $L_{j-i}(\max(0,j-i),T-i)$. For $T-i\geq j$ the difference is

$$\sum_{h=1+\max(0,j-i)}^{T-i} \iota_h\iota_{h+i-j}^T - \sum_{h=1+\max(0,j-i)}^{j} \iota_h\iota_{h+i-j}^T = \sum_{h=j+1}^{T-i} \iota_h\iota_{h+i-j}^T = L_{j-i}(j). \ \square$$

Proposition 3.5

The differential of the inverse of the AR covariance matrix is

$$\mathbf{d}V_{AR}^{-1} = \sum_{i=1}^{p}\sum_{j=0}^{p}(L_{j-i}(j) + L_{j-i}^T(j))\vartheta_j\mathbf{d}\vartheta_i.$$

Proof

From proposition 3.4 we conclude

$$\mathbf{d}V_{AR}^{-1} = \sum_{i=0}^{p}\sum_{j=0}^{p} L_{j-i}(j)(\vartheta_j\mathbf{d}\vartheta_i + \vartheta_i\mathbf{d}\vartheta_j).$$

Writing out and interchanging the indices in the second part gives

$$\mathbf{d}V_{AR}^{-1} = \sum_{i=0}^{p}\sum_{j=0}^{p} L_{j-i}(j)\vartheta_j\mathbf{d}\vartheta_i + \sum_{i=0}^{p}\sum_{j=0}^{p} L_{i-j}(i)\vartheta_j\mathbf{d}\vartheta_i$$

$$= \sum_{i=1}^{p}\sum_{j=0}^{p}(L_{j-i}(j) + L_{j-i}^T(j))\vartheta_j\mathbf{d}\vartheta_i, \text{ as } \mathbf{d}\vartheta_0 = 0. \ \square$$

Proposition 3.6

Let $\underline{V}_{AR}^{-1} = \underline{P}^T\underline{P}-\underline{Q}\underline{Q}^T$, the upper left $(p\times p)$ submatrix of the inverse of the AR covariance matrix. Then

$$\underline{V}_{AR}^{-1} = \sum_{i=0}^{p}\Big(\sum_{j=0}^{p-i-1} L_{j-i}(j)\vartheta_j - \sum_{j=p-i+1}^{p} L_{j-i}(p-i)\vartheta_j\Big)\vartheta_i.$$

Proof

Using lag matrices of order $p\times p$ we get $\underline{P} = \sum_{j=0}^{p-1} L_j(j)\vartheta_j$ and $\underline{Q} = \sum_{j=1}^{p} L_{j-p}\vartheta_j$, but as $L_p(p)\vartheta_p = L_{-p}\vartheta_0 = 0$, we write

$$\underline{V}_{AR}^{-1} = \Big(\sum_{i=0}^{p} L_i(i)\vartheta_i\Big)^T \sum_{j=0}^{p} L_j(j)\vartheta_j - \Big(\sum_{i=0}^{p} L_{i-p}\vartheta_i\Big)\Big(\sum_{j=0}^{p} L_{j-p}\vartheta_j\Big)^T.$$

$$= \sum_{i=0}^{p} \sum_{j=0}^{p} L_{-i}L_j(j)\vartheta_i\vartheta_j - \sum_{i=0}^{p} \sum_{j=0}^{p} L_{i-p}L_{p-j}(p-j)\vartheta_i\vartheta_j$$

$$= \sum_{i=0}^{p} \sum_{j=0}^{p} L_{j-i}(\max(0,j-i),j)\vartheta_i\vartheta_j - \sum_{i=0}^{p} \sum_{j=0}^{p} L_{i-j}(\max(0,-j+i),p-j)\vartheta_i\vartheta_j$$

$$= \sum_{i=0}^{p} \sum_{j=0}^{p} (L_{j-i}(\max(0,j-i),j) - L_{j-i}(\max(0,j-i),p-i))\vartheta_i\vartheta_j.$$

Conforming the definition we get for the lag matrices:

$L_{j-i}(\max(0,j-i),j) - L_{j-i}(\max(0,j-i),p-i) =$

$$= \sum_{h=1+\max(0,j-i)}^{p-i} \iota_h \iota_{h+i-j}^T - \sum_{h=1+\max(0,j-i)}^{j} \iota_h \iota_{h+i-j}^T$$

$$= \begin{cases} \sum_{h=j+1}^{p-i} \iota_h \iota_{h+i-j}^T = L_{j-i}(j) & \text{if } j+i<p \\ 0 & \text{if } j+i=p \\ -\sum_{h=p-i+1}^{j} \iota_h \iota_{h+i-j}^T = -L_{j-i}(p-i) & \text{if } j+i>p \end{cases}$$

Hence,

$$\underline{V}_{AR}^{-1} = \sum_{i=0}^{p} \Big(\sum_{j=0}^{p-i-1} L_{j-i}(j)\vartheta_j - \sum_{j=p-i+1}^{p} L_{j-i}(p-i)\vartheta_j \Big)\vartheta_i. \quad \square$$

Remark. As $\underline{V}$ is symmetric we can also write:

$$\underline{V}_{AR}^{-1} = \sum_{i=0}^{p} \Big(\sum_{j=0}^{p-i-1} L_{j-i}^T(j)\vartheta_j - \sum_{j=p-i+1}^{p} L_{j-i}^T(p-i)\vartheta_j \Big)\vartheta_i$$

$$= \sum_{i=0}^{p} \Big(\sum_{j=0}^{p-i-1} L_{i-j}(i)\vartheta_j - \sum_{j=p-i+1}^{p} L_{i-j}(p-j)\vartheta_j \Big)\vartheta_i$$

$$= \sum_{j=0}^{p} \Big(\sum_{i=0}^{p-j-1} L_{j-i}(j)\vartheta_i - \sum_{i=p-j+1}^{p} L_{j-i}(p-i)\vartheta_i \Big)\vartheta_j. \quad \square$$

For the differential $\mathbf{d}\underline{V}_{AR}^{-1}$ we have

Proposition 3.7

$$\mathbf{d}\underline{V}_{AR}^{-1} = \sum_{i=1}^{p} \Big(\sum_{j=0}^{p-i-1} (L_{j-i}(j) + L_{j-i}^{T}(j))\vartheta_j - \sum_{j=p-i+1}^{p} (L_{j-i}(p-i) + L_{j-i}^{T}(p-i))\vartheta_j \Big) \mathbf{d}\vartheta_i.$$

Proof

$$\mathbf{d}\underline{V}_{AR}^{-1} = \sum_{i=0}^{p} \Big(\sum_{j=0}^{p-i-1} L_{j-i}(j)\vartheta_j - \sum_{j=p-i+1}^{p} L_{j-i}(p-i)\vartheta_j \Big) \mathbf{d}\vartheta_i +$$

$$\sum_{j=0}^{p} \Big(\sum_{i=0}^{p-j-1} L_{j-i}(j)\vartheta_i - \sum_{i=p-j+1}^{p} L_{j-i}(p-\}i)\vartheta_i \Big) \mathbf{d}\vartheta_j.$$

Interchanging i and j in the second part we get

$$\mathbf{d}\underline{V}_{AR}^{-1} = \sum_{i=0}^{p} \Big(\sum_{j=0}^{p-i-1} L_{j-i}(j)\vartheta_j - \sum_{j=p-i+1}^{p} L_{j-i}(p-i)\vartheta_j \Big) \mathbf{d}\vartheta_i +$$

$$\sum_{i=0}^{p} \Big(\sum_{j=0}^{p-i-1} L_{i-j}(i)\vartheta_j - \sum_{j=p-i+1}^{p} L_{i-j}(p-j)\vartheta_j \Big) \mathbf{d}\vartheta_i.$$

or as $L_{i-j}(i) = L_{j-i}^{T}(j)$ and $\mathbf{d}\vartheta_0 = 0$

$$\mathbf{d}\underline{V}_{AR}^{-1} = \sum_{i=1}^{p} \Big(\sum_{j=0}^{p-i-1} (L_{j-i}(j) + L_{j-i}^{T}(j))\vartheta_j - \sum_{j=p-i+1}^{p} (L_{j-i}(p-i) + L_{j-i}^{T}(p-i))\vartheta_j \Big) \mathbf{d}\vartheta_i. \;\square$$

Moreover we need the lag form of P^TP and MM^T for CLS estimation. The following propositions treat both expressions and their differentials.

Proposition 3.8

$$P^TP = \sum_{i=0}^{p} \Big(\sum_{j=0}^{i} L_{j-i}(0,j)\vartheta_j + \sum_{j=i+1}^{p} L_{i-j}^{T}(0,i)\vartheta_j \Big) \vartheta_i.$$

Proof

Applying the expression for the lag form of P, taking the transpose and multiplying gives

$$P^TP = \Big(\sum_{i=0}^{p} L_i(i)\vartheta_i\Big)^T \sum_{j=0}^{p} L_j(j)\vartheta_j$$

$$= \sum_{i=0}^{p}\ \sum_{j=0}^{p} L_i^T(i)L_j(j)\vartheta_j\vartheta_i$$

$$= \sum_{i=0}^{p}\ \sum_{j=0}^{p} L_{-i}L_j(j)\vartheta_j\vartheta_i$$

$$= \sum_{i=0}^{p}\ \sum_{j=0}^{p} L_{j-i}(\max(0,j\text{-}i),j)\vartheta_j\vartheta_i$$

$$= \sum_{i=0}^{p}\Big(\sum_{j=0}^{i} L_{j-i}(0,j)\vartheta_j + \sum_{j=i+1}^{p} L_{j-i}(j\text{-}i,j)\vartheta_j\Big)\vartheta_i$$

$$= \sum_{i=0}^{p}\Big(\sum_{j=0}^{i} L_{j-i}(0,j)\vartheta_j + \sum_{j=i+1}^{p} L_{i-j}^T(0,i)\vartheta_j\Big)\vartheta_i. \quad \square$$

Proposition 3.9

$$\mathbf{d}(P^TP) = \sum_{i=0}^{p}\Big\{\sum_{j=0}^{i}(L_{j-i}(0,j) + L_{j-i}^T(0,j)) + \sum_{j=i+1}^{p}(L_{i-j}(0,i) + L_{i-j}^T(0,i))\vartheta_j\Big\}\mathbf{d}\vartheta_i.$$

Proof

The differential can be obtained from proposition 3.8, but it is easier to derive it directly. For the differential we get

$$\mathbf{d}(P^TP) = P^T\mathbf{d}P + \mathbf{d}(P^T)P$$

$$= \sum_{i=0}^{p} L_i^T(i)\vartheta_i \sum_{j=0}^{p} L_j(j)\mathbf{d}\vartheta_j + \sum_{i=0}^{p} L_i^T(i)\mathbf{d}\vartheta_i \sum_{j=0}^{p} L_j(j)\vartheta_j$$

$$= \sum_{i=0}^{p}\ \sum_{j=0}^{p} L_{-i}L_j(j)\vartheta_i\mathbf{d}\vartheta_j + \sum_{i=0}^{p}\ \sum_{j=0}^{p} L_{-i}L_j(j)\vartheta_j\mathbf{d}\vartheta_i$$

$$= \sum_{i=0}^{p}\ \sum_{j=0}^{p} L_{j-i}(\max(0,j\text{-}i),j)\vartheta_i\mathbf{d}\vartheta_j + \sum_{i=0}^{p}\ \sum_{j=0}^{p} L_{j-i}(\max(0,j\text{-}i),j)\vartheta_j\mathbf{d}\vartheta_i$$

$$= \sum_{i=0}^{p}\ \sum_{j=0}^{p} L_{i-j}(\max(0,i\text{-}j),i)\vartheta_j\mathbf{d}\vartheta_i + \sum_{i=0}^{p}\ \sum_{j=0}^{p} L_{j-i}(\max(0,j\text{-}i),j)\vartheta_j\mathbf{d}\vartheta_i$$

$$= \sum_{i=0}^{p} \{ \sum_{j=0}^{i} (L_{i-j}(i\text{-}j,i) + L_{j-i}(0,j)\vartheta_j) + \sum_{j=i+1}^{p} (L_{i-j}(0,i) + L_{j-i}(j\text{-}i,j)\vartheta_j\} \mathbf{d}\vartheta_i +$$

$$= \sum_{i=0}^{p} \{ \sum_{j=0}^{i} (L^{T}_{j-i}(0,j) + L_{j-i}(0,j)\}\vartheta_j + \sum_{j=i+1}^{p} (L_{i-j}(0,i) + L^{T}_{i-j}(0,i))\vartheta_j\} \mathbf{d}\vartheta_i. \ \square$$

Proposition 3.10

$$MM^{T} = \sum_{i=0}^{q} L_i(i)\alpha_i \sum_{j=0}^{q} L^{T}_j(j)\alpha_j.$$

Proof

The proof follows the same lines as in the case of the AR parameters. However, there is one difference, as the transposed part comes first in the AR case. Hence the result is slightly different.

M is equal to $\sum_{i=0}^{q} L_i(i)\alpha_i$, which gives for MM^T:

$$MM^{T} = \sum_{i=0}^{q} L_i(i)\alpha_i \sum_{j=0}^{q} L^{T}_j(j)\alpha_j$$

$$= \sum_{i=0}^{q} \sum_{j=0}^{q} L_i(i)L_{-j}\alpha_i\alpha_j$$

$$= \sum_{i=0}^{q} \sum_{j=0}^{q} L_{i-j}(i,\max(i\text{-}j,0))\alpha_i\alpha_j$$

$$= \sum_{i=0}^{q} \{ \sum_{j=0}^{i} L_{i-j}(i,i\text{-}j)\alpha_j + \sum_{j=i+1}^{q} L_{i-j}(i,0)\alpha_j\}\alpha_i$$

$$= \sum_{i=0}^{q} \{ \sum_{j=0}^{i} L^{T}_{j-i}(j,0)\alpha_j + \sum_{j=i+1}^{q} L_{i-j}(i,0)\alpha_j\}\alpha_i. \ \square$$

Proposition 3.11

$$\mathbf{d}(MM^{T}) = \sum_{i=0}^{q} \{ \sum_{j=0}^{i} L^{T}_{j-i}(j,0) + L_{j-i}(j,0)\alpha_j + \sum_{j=i+1}^{q} L_{i-j}(i,0) + L^{T}_{i-j}(i,0)\alpha_j\} \mathbf{d}\alpha_i.$$

Proof

$$\mathbf{d}(MM^T) = (\mathbf{d}M)M^T + M(\mathbf{d}M^T)$$

$$= \sum_{i=0}^{q} L_i(i)\mathbf{d}\alpha_i \sum_{j=0}^{q} L_j^T(j)\alpha_j + \sum_{i=0}^{q} L_i(i)\alpha_i \sum_{j=0}^{q} L_j^T(j)\mathbf{d}\alpha_j$$

$$= \sum_{i=0}^{q} \sum_{j=0}^{q} L_i(i)L_{-j}\alpha_j\mathbf{d}\alpha_i + \sum_{i=0}^{q} \sum_{j=0}^{q} L_i(i)L_{-j}\alpha_i\mathbf{d}\alpha_j$$

$$= \sum_{i=0}^{q} \sum_{j=0}^{q} L_{i-j}(i,\max(i-j,0))\alpha_j\mathbf{d}\alpha_i + \sum_{i=0}^{q} \sum_{j=0}^{q} L_{i-j}(i,\max(i-j,0))\alpha_i\mathbf{d}\alpha_j$$

$$= \sum_{i=0}^{q} \sum_{j=0}^{q} L_{i-j}(i,\max(i-j,0))\alpha_j\mathbf{d}\alpha_i + L_{j-i}(j,\max(j-i,0))\alpha_j\mathbf{d}\alpha_i$$

$$= \sum_{i=0}^{q} \{ \sum_{j=0}^{i} (L_{i-j}(i,i-j) + L_{j-i}(j,0))\alpha_j + \sum_{j=i+1}^{q} (L_{i-j}(i,0) + L_{j-i}(j,j-i))\alpha_j\}\mathbf{d}\alpha_i$$

$$= \sum_{i=0}^{q} \{ \sum_{j=0}^{i} L_{j-i}^T(j,0) + L_{j-i}(j,0)\alpha_j + \sum_{j=i+1}^{q} L_{i-j}(i,0) + L_{i-j}^T(i,0)\alpha_j\}\mathbf{d}\alpha_i. \quad \square$$

3.3 Summary

In this chapter a so called lag matrix was defined. In fact the definition is not restricted to lags, but also leads can be handled. Furthermore some useful properties regarding transposition and multiplication are presented. This lag matrix is used to write the covariance matrix, or parts of it, in a form that is suited for differentiating. Most of the propositions are rather technical and will be used in the following chapters.

IV DERIVATIVES OF THE LIKELIHOOD FUNCTION

4.1 Introduction

Several authors have given procedures to estimate the parameters of an ARMA model, either for a pure time series model (Kohn and Ansley, 1985) or for a regression model (Zinde-Walsh and Galbraith, 1991). Kohn and Ainsley state *we gain efficiency by working with a MA(q) rather than an ARMA(p,q) model for the first N-p observations*, where N is the sample size. One can wonder whether such a procedure is very useful. Zinde-Walsh and Galbraith also use an approximation: they write *that solving equations in the first order conditions is not much different from solving an equation which ignores the determinant of the covariance matrix* (p. 338). But even with several simplifications the resulting formulas and algorithms to estimate the ARMA parameters become complicated without a closed form of the general ARMA covariance matrix. Given the Aitken estimator and the corresponding residuals we derive first derivatives of the likelihood function, as defined in (1.12), of the ARMA-parameters.
At the same time we are able to show the differences between maximum likelihood and minimum distance estimators. In the case of a pure AR and MA model the difference amounts to a sum of the elements of the off-diagonals of the covariance matrix. The general ARMA model has the same property, but the function of the elements of the covariance matrix is more complicated.

Our aim is to maximise the loglikelihood, or equivalently to minimise (1.12). This is the same as minimising $S = \log|V| + T\cdot\log e^{T}V^{-1}e$, which we will call the modified likelihood function. The reason to use this expression is the fact that it differentials have less terms. For the differential of the determinant part we use $\mathbf{d}(|V|) = |V|\,\mathbf{tr}V^{-1}\mathbf{d}V$ (see *e.g.* Magnus and Neudecker, 1988, p.149), while the differential of the quadratic form is $e^{T}\mathbf{d}V^{-1}e$.
The result is the first differential

$$\mathbf{d}S = \mathrm{tr}V^{-1}\mathbf{d}V + \frac{1}{s^2}e^T\mathbf{d}(V^{-1})e \tag{4.1}$$

with $s^2 = e^TV^{-1}e/T$. This, in a slightly different form, is what Magnus (1978) called the ϑ-equation(s).

In the sequel we will show how ϑ and α can expressed as a function of e, be it computed or identical to y. The resulting first order conditions can be written as a linear function of the parameter to be estimated, say $A\xi + b$, with $\xi = \alpha$ in the MA case and $\xi = \vartheta$ in the AR case, where A is a matrix and b a vector both depending on V. However it is not possible to find a simple solution in the form $\xi = A^{-1}b$. Thus we have to employ other methods to find a minimum. The results here give us an analytical expression for the first derivatives of $\partial S/\partial\vartheta$ and $\partial S/\partial\alpha$ which are useful to find solutions.

4.2 First derivatives

From (4.1) we derive the first derivatives of the loglikelihood. We make use of the lag form for the covariance matrix as given in chapter 3. First we present a theorem for the general ARMA case, next we give two corollaries for the AR and the MA case.
As $V_{i,i+k} = V_{i+k,i}$ for $k=0,.., T-1$ and $1 \le i \le T-k$, we define $V[k] = V_{i,i+k}$, the determining element of the k^{th} (off-)diagonal.

Theorem 4.1

Let $\frac{\partial S}{\partial\vartheta} = \left(\frac{\partial S}{\partial\vartheta_1}, \frac{\partial S}{\partial\vartheta_2}, \dots, \frac{\partial S}{\partial\vartheta_p}\right)^T$ be the vector of AR derivatives and $\frac{\partial S}{\partial\alpha} = \left(\frac{\partial S}{\partial\alpha_1}, \frac{\partial S}{\partial\alpha_2}, \dots, \frac{\partial S}{\partial\alpha_q}\right)^T$ the vector of MA derivatives.

AR-part

Let $Z = \bar{V}_{AR}[N\ M]^TV^{-1}$, $Z \in \mathbb{R}^{(T+q)xT}$, $\varsigma = Ze \in \mathbb{R}^{(T+q)x1}$ and $\psi(i,j)$ be the sum of the elements of $(i\text{-}j)^{th}$ diagonal of ZVZ^T, without the first and last min(i,j) elements. Then $\frac{\partial S}{\partial \vartheta} = G\vartheta + g$, where the $(i,j)^{th}$ element of $G \in \mathbb{R}^{p\times p}$ is

$$g_{i,j} = 2\{\sum_{k=1}^{T+q-i-j} \zeta_{k+i}\zeta_{k+j}/s^2 - \psi(i,j)\} \tag{4.2}$$

and the i^{th} element of $g \in \mathbb{R}^{p\times 1}$ is $g_i = g_{i,0}$.

MA-part

Let $\phi = V^{-1}e$. Then $\frac{\partial S}{\partial \alpha} = H\alpha + h$, where the $(i,j)^{th}$ element of $H \in \mathbb{R}^{q\times q}$ is

$$h_{i,j} = -2\sum_{l=1}^{T}\sum_{k=1}^{T} \bar{V}_{AR}[l-k-i+j](\phi_l\phi_k/s^2 - V^{-1}[k,l]) \tag{4.3}$$

and the i^{th} element of $h \in \mathbb{R}^{q\times 1}$ is $h_i = h_{i,0}$.

Proof

AR-part

We have to evaluate $\mathbf{tr}V^{-1}\mathbf{d}V$ and $e^T\mathbf{d}V^{-1}e$. However it is not possible to isolate the AR-parameters in the expression $V^{-1} = ([N\ M]\bar{V}_{AR}[N\ M]^T)^{-1}$, where $\bar{V}_{AR} = [\bar{P}^T\bar{P} - \bar{Q}\bar{Q}^T]^{-1}$. A way out is to define $Z = \bar{V}_{AR}[N\ M]^T V^{-1}$, which implies $V^{-1} = Z^T\bar{V}_{AR}^{-1}Z$, as is easily verified. The differential in the ϑ-direction is

$$\begin{aligned}\mathbf{d}V^{-1} &= -V^{-1}\mathbf{d}VV^{-1}\\ &= -V^{-1}[N\ M]\mathbf{d}\bar{V}_{AR}[N\ M]^T V^{-1}\\ &= Z^T\mathbf{d}\bar{V}_{AR}^{-1}Z.\end{aligned}$$

The quadratic form $e^T\mathbf{d}V^{-1}e$ becomes, using $\zeta = Ze$ and proposition 3.5,

$$\begin{aligned}e^T\mathbf{d}V^{-1}e &= \zeta^T\mathbf{d}\bar{V}_{AR}^{-1}\zeta\\ &= \zeta^T(\sum_{i=1}^{p}\sum_{j=0}^{p}(L_{j-i}(j) + L_{j-i}^T(j))\vartheta_j\mathbf{d}\vartheta_i)\zeta\\ &= 2\sum_{i=1}^{p}\sum_{j=0}^{p}\zeta^T L_{j-i}(j)\zeta\ \vartheta_j\mathbf{d}\vartheta_i\\ &= 2\sum_{i=1}^{p}\sum_{j=0}^{p}\sum_{h=1}^{T+q-i-j}\zeta_{h+i}\zeta_{h+j}\vartheta_j\mathbf{d}\vartheta_i.\end{aligned}$$

Here $\zeta^T L_{j-i}(j)\zeta = \sum_{h=1+j}^{T+q-i}\zeta^T\iota_h\iota_{h-j+i}^T\zeta = \sum_{h=1}^{T+q-i-j}\zeta_{h+j}\zeta_{h+i}$.

Hence $e^T\partial(V^{-1})e/\partial\vartheta_i = 2\sum_{j=0}^{p}\sum_{h=1}^{T+q-i-j} \zeta_{h+i}\zeta_{h+j}\vartheta_j$, $i=1,..,p$.

The determinantal part becomes, using lag matrices and some basic properties of the trace operator:

$\mathbf{tr}V^{-1}\mathbf{d}V = -\mathbf{tr}V\mathbf{d}V^{-1} = -\mathbf{tr}VZ^T\mathbf{d}\bar{V}_{AR}^{-1}Z =$

$$= -\mathbf{tr}ZVZ^T(\sum_{i=0}^{p}\sum_{j=0}^{p}(L_{j-i}(j)+L_{j-i}^T(j))\vartheta_j\mathbf{d}\vartheta_i)$$

$$= -2\sum_{i=1}^{p}\sum_{j=0}^{p}\mathbf{tr}ZVZ^TL_{j-i}(j)\vartheta_j\mathbf{d}\vartheta_i$$

$$= -2\sum_{i=1}^{p}\sum_{j=0}^{p}\mathbf{tr}ZVZ^T\sum_{h=1+j}^{T+q-i}\iota_h\iota_{h-j+i}^T\vartheta_j\mathbf{d}\vartheta_i$$

$$= -2\sum_{i=1}^{p}\sum_{j=0}^{p}\sum_{h=1}^{T+q-i-j}\iota_{h+j}^TZVZ^T\iota_{h+i}\vartheta_j\mathbf{d}\vartheta_i.$$

Here $\psi(i,j) = \sum_{h=1}^{T+q-i-j}\iota_{h+j}^TZVZ^T\iota_{h+i}$ is the sum of the elements of $(i-j)^{th}$ diagonal of ZVZ^T without the first and last min(i,j) elements, or

$\mathbf{tr}V^{-1}\partial V/\partial\vartheta_i = -2\sum_{j=0}^{p}\sum_{h=1}^{T+q-i-j}\iota_{h+j}^TZVZ^T\iota_{h+i}\vartheta_j$, $i=1,..,p$.

End of proof AR-part.

MA-part

From proposition 3.3 we have

$$\mathbf{d}V = \sum_{i=1}^{q}\sum_{j=0}^{q}\sum_{h=1}^{T}\sum_{k=1}^{T}\bar{V}_{AR}[h-k-i+j](\iota_h\iota_k^T+\iota_k\iota_h^T)\alpha_j\mathbf{d}\alpha_i.$$

The quadratic part becomes

$e^T\mathbf{d}V^{-1}e = -e^TV^{-1}(\mathbf{d}V)V^{-1}e =$

$$= -\phi^T\{\sum_{i=1}^{q}\sum_{j=0}^{q}\sum_{h=1}^{T}\sum_{k=1}^{T}\bar{V}_{AR}[h-k-i+j](\iota_h\iota_k^T+\iota_k\iota_h^T)\alpha_j\mathbf{d}\alpha_i\}\phi$$

$$=-\sum_{i=1}^{q}\sum_{j=0}^{q}\sum_{h=1}^{T}\sum_{k=1}^{T}\bar{V}_{AR}[h-k-i+j](\phi^T\iota_h\iota_k^T\phi+\phi^T\iota_k\iota_h^T\phi)\alpha_j\mathbf{d}\alpha_i$$

$$=-2\sum_{i=1}^{q}\sum_{j=0}^{q}\sum_{h=1}^{T}\sum_{k=1}^{T}\bar{V}_{AR}[h-k-i+j]\phi_h\phi_k\alpha_j\mathbf{d}\alpha_i$$

or $e^T\partial(V^{-1})e/\partial\alpha_i=-2\sum_{j=0}^{q}\sum_{h=1}^{T}\sum_{k=1}^{T}\bar{V}_{AR}[h-k-i+j]\phi_h\phi_k\alpha_j,\ i=1,..,q.$

For the determinantal part we get, using proposition 3.3:

$$\mathbf{tr}V^{-1}\mathbf{d}V=\mathbf{tr}\{V^{-1}\sum_{i=1}^{q}\sum_{j=0}^{q}\sum_{h=1}^{T}\sum_{k=1}^{T}\bar{V}_{AR}[h-k-i+j](\iota_h\iota_k^T+\iota_k\iota_h^T)\alpha_j\mathbf{d}\alpha_i\}$$

$$=\sum_{i=1}^{q}\sum_{j=0}^{q}\sum_{h=1}^{T}\sum_{k=1}^{T}\bar{V}_{AR}[h-k-i+j](\iota_k^TV^{-1}\iota_h+\iota_h^TV^{-1}\iota_k)\alpha_j\mathbf{d}\alpha_i$$

$$=2\sum_{i=1}^{q}\sum_{j=0}^{q}\sum_{h=1}^{T}\sum_{k=1}^{T}\bar{V}_{AR}[h-k-i+j]V^{-1}[k,h]\alpha_j\mathbf{d}\alpha_i$$

or $\mathbf{tr}V^{-1}\partial V/\partial\vartheta_i=\ 2\sum_{j=0}^{q}\sum_{h=1}^{T}\sum_{k=1}^{T}\bar{V}_{AR}[h-k-i+j]V^{-1}[k,h]\alpha_j,\ i=1,..,q.$

End of proof MA-part. □

Corollary 4.1

Derivatives of the AR likelihood function

$\frac{\partial S}{\partial\vartheta}=\ G\vartheta+g$, where the $(i,j)^{th}$ element of G is

$$g_{i,j}=2\{\sum_{k=1}^{T-i-j}e_{k+i}e_{k+j}/s^2-(p-i-j)\underline{V}_{AR}[i-j]\} \tag{4.4}$$

and the i^{th} element of $g\in\mathbb{R}^{p\times 1}$ is $g_i=g_{i,0}$.

Proof

For $q=0$, $\bar{V}_{AR}=V_{AR}=(P^TP-QQ^T)^{-1}$, $[N\ \ M]=I_T$ and $V=V_{AR}$, which gives $Z=I_T$ or $\zeta=e$. For the determinant part we use $|V_{AR}|=|\underline{V}_{AR}|$ and proposition 3.7: $\mathbf{tr}V^{-1}\partial V/\partial\vartheta_i=-\mathbf{tr}V\partial V^{-1}/\partial\vartheta_i=-\mathbf{tr}\underline{V}_{AR}\partial\underline{V}_{AR}^{-1}/\partial\vartheta_i=$

$$=-\mathbf{tr}\underline{V}_{AR}\{\sum_{j=0}^{p-i-1}(L_{j-i}(j)+L^T_{j-i}(j))\vartheta_j-\sum_{j=p-i+1}^{p}(L_{j-i}(p-i)+L^T_{j-i}(p-i))\vartheta_j)\}.$$

Because $\mathbf{tr}\underline{V}_{AR}L^T_k(m)=\mathbf{tr}(\underline{V}_{AR}L^T_k(m))^T=\mathbf{tr}L_k(m)\underline{V}_{AR}=\mathbf{tr}\underline{V}_{AR}L_k(m)=$

$\mathbf{tr}\sum_{h=1+m}^{p+k-m}\underline{V}_{AR}\iota_h\iota^T_{h-k}=\sum_{h=1+m}^{p+k-m}\underline{V}_{AR}[k]=(p+k-2m)\underline{V}_{AR}[k]$ we get

$$\mathbf{tr}V^{-1}\partial V/\partial\vartheta_i=-2\{\sum_{j=0}^{p-i-1}(p+j-i-2j)\underline{V}_{AR}[j-i]\vartheta_j-\sum_{j=p-i+1}^{p}(p+j-i-2(p-i))\underline{V}_{AR}[j-i]\vartheta_j)\}$$

$$=-2\{\sum_{j=0}^{p-i-1}(p-j-i)\underline{V}_{AR}[j-i]\vartheta_j-\sum_{j=p-i+1}^{p}(-p+j+i)\underline{V}_{AR}[j-i]\vartheta_j)\}$$

$$=-2\sum_{j=0}^{p}(p-j-i)\underline{V}_{AR}[j-i]\vartheta_j.\ \square$$

As an illustration we will show how the well known cubic equation for the AR(1) case as published by Beach and MacKinnon (1978) can be found using this corollary. They stated:

"$f(\rho)=\rho^3+a\rho^2+b\rho+c=0$, where
$a=-(T-2)\Sigma e_te_{t-1}/[(T-1)(\Sigma e^2_{t-1}-e^2_1)]$
$b=[(T-1)e^2_1-T\Sigma e^2_{t-1}-\Sigma e^2_t]/[(T-1)(\Sigma e^2_{t-1}-e^2_1)]$
$c=T\Sigma e_te_{t-1}/[(T-1)(\Sigma e^2_{t-1}-e^2_1)],...$; the summations run from $t=2$ to T."

Define $c_1=\sum_{t=2}^{T-1}e^2_t$, $c_2=\sum_{t=1}^{T-1}e_te_{t+1}$ and $c_3=\sum_{t=1}^{T}e^2_t$, replace ρ by $-\vartheta$ and their cubic equation becomes $(T-1)c_1\vartheta^3-(T-2)c_2\vartheta^2-(Tc_1+c_2)\vartheta+Tc_2=0$. Using corollary 4.1 with $p=1$ we get $\underline{P}=1$, $\underline{Q}=\vartheta$ and $\underline{V}_{AR}[0]=(\underline{P}^T\underline{P}-\underline{QQ}^T)^{-1}=(1-\vartheta^2)^{-1}$. The elements of G become

$$g_{1,1}=2\{\sum_{k=1}^{T-1-1}e_{k+1}e_{k+1}/s^2-(1-1-1)\underline{V}_{AR}[1-1]\}=\frac{2}{s^2}\{c_1+\frac{s^2}{1-\vartheta^2}\}\text{ and}$$

$$g_{1,0}=2\{\sum_{k=1}^{T-1-0}e_{k+1}e_{k+0}/s^2-(1-1-0)\underline{V}_{AR}[1-0]\}=2\frac{c_2}{s^2}.$$

Observe that $T\cdot s^2=e^TV^{-1}e=e^T(P^TP-QQ^T)e=(Pe)^T(Pe)-(Q^Te)^T(Q^Te)$ with

$Pe=(e_1\ \ \vartheta e_1+e_2\ \dots\ \vartheta e_{T-1}+e_T)^T$ and $Q^Te=(\vartheta e_1\ \ 0\ \dots\ 0)^T$, which gives $s^2=(c_1\vartheta^2+2c_2\vartheta+c_3)/T$.

This results in $c_1\vartheta + \dfrac{c_1\vartheta^2+2c_2\vartheta+c_3}{T}\dfrac{\vartheta}{1-\vartheta^2} - c_2=0$, which gives after some manipulations the pseudo cubic relation of Beach and MacKinnon. The same result can also be found in Magnus (1978).

Corollary 4.2

Derivatives of the MA likelihood function

Let $\phi=V_{MA}^{-1}$ and d_k the sum of the elements of the k^{th} off-diagonal of V_{MA}^{-1}. Then

$\dfrac{\partial S}{\partial\alpha}= H\alpha+h$, where the $(i,j)^{th}$ element of H is

$$h_{i,j}= 2\{\sum_{h=1}^{T-|i-j|}\phi_h\phi_{h+|i-j|}/s^2- d_{|i-j|}\}\tag{4.5}$$

and the i^{th} element of $h\in\mathbb{R}^{q\times1}$ is $h_i=h_{i,0}$.

Proof

Starting from (4.3) we have $\bar{V}_{AR}=I_{T+q}$. Hence $\bar{V}_{AR}[h-k-i+j]$ is equal to 1 if and only if $h-k-i+j=0$, or $k=h-i+j$ and zero elsewhere. Put $g=h+\min(0,j-i)$, then $k=h-i+j=g-\min(i-j,0)$ and g runs from $\max(1+\min(0,j-i),\ 1+\min(i-j,0))$ to $\min(T+\min(0,j-i),\ \min(0,i-j))$ or from 1 to $T-|i-j|$. The expression $\phi_h\phi_k$ becomes $\phi_{g-\min(0,j-i)}\phi_{g-\min(0,i-j)}=\phi_{g+\max(0,i-j)}\phi_{g+\max(0,j-i)}= \phi_g\phi_{g+|i-j|}$.

In the same way we get for $\sum_{h=1}^{T}\sum_{k=1}^{T}\bar{V}_{AR}[h-k-i+j]V^{-1}[k,h]=$

$\sum_{g=1}^{T-|i-j|}V_{MA}^{-1}[g,g+|i-j|]=d_{|i-j|}$. □

Computational remarks

Without going into the details we give an outline how the derivatives can be found using the elementary matrices (M, N, P and Q) and several expression derived from them. For both the AR and MA derivatives we need e, the difference between y and Xb, with $b=(X^TV^{-1}X)^{-1}X^TV^{-1}y$ and $s^2=e^TV^{-1}e/T$. While the inverse of V is necessary for the computations it

is easy though not necessary to have the disposal of V itself. Furthermore we need $\bar{V}_{AR}$, the AR covariance matrix of order $(T+r)x(T+r)$.
The inverse of V can be found using (2.7) which contains $R(=M^{-1}PN-Q) \in \mathbb{R}^{Txr}$ and $\underline{D}(=\underline{P}^T\underline{P}-\underline{Q}\underline{Q}^T) \in \mathbb{R}^{rxr}$, defined in lemma (2.4). The matrices V^{-1} and R contain $M^{-1}P$, which is a lower band matrix: as M is a lower band matrix, its inverse is of the same type. The inverse of M can be obtained by a simple recursive algorithm and the numerical values of its elements tend to zero: they are in general closer to zero the larger their distance to the main diagonal because of the invertibility condition. As $\underline{D}$ is of order rxr its inverse is easily computed. Hence V^{-1} can be found by straightforward matrix calculus.
The expression for $\bar{V}_{AR}$ is already treated in the proof of lemma (2.2). In partitioned form it gives no problems; $\underline{D}^{-1}$ is as above and for P^{-1} the same remarks apply as for M^{-1}. For V we use (2,4), of which all elements are already treated here. We conclude that there is no need to invert large matrices. What remains is the computation of Z, ZVZ^T, ς and ϕ. The elements of H, h, G and g follow directly.

4.3 Second differential

In this section we will derive an expression for the second differential of the likelihood function as the first step to obtain the second derivatives of the loglikelihood. The second derivative itself is helpful in finding the minimum of the modified likelihood function, which is quite complicated. It gives us the direction in which a local minimum can be found. Secondly, from the expression of the second derivatives we can conclude that in general a stationary point need not to be a minimum, as we can never conclude that is has to be positive. A third reason is to find an expression for the information matrix, which is part of the second derivative.

Before we differentiate (4.1) again we first derive the differential of $e = y - Xb$:

$$\begin{aligned}\mathbf{d}e &= -X\mathbf{d}b = -X\mathbf{d}((X^TV^{-1}X)^{-1}X^TV^{-1}y)\\ &= X(X^TV^{-1}X)^{-1}X^T\mathbf{d}V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}y - X(X^TV^{-1}X)^{-1}X^T\mathbf{d}V^{-1}y\\ &= -X(X^TV^{-1}X)^{-1}X^T\mathbf{d}V^{-1}e.\end{aligned}$$

Hence $e^T\mathbf{d}V^{-1}\mathbf{d}e = -e^T\mathbf{d}V^{-1}X(X^TV^{-1}X)^{-1}X^T\mathbf{d}V^{-1}e$, which is clearly negative definite.

Straightforward differentiation of (4.1) gives for the second differential:

$$\begin{aligned}\mathbf{d}^2S &= \mathbf{d}\{\mathbf{tr}V^{-1}\mathbf{d}V + T(e^TV^{-1}e)^{-1}e^T\mathbf{d}V^{-1}e\}\\ &= \mathbf{tr}\mathbf{d}V^{-1}\mathbf{d}V + \mathbf{tr}V^{-1}\mathbf{d}^2V + T\mathbf{d}(e^TV^{-1}e)^{-1}e^T\mathbf{d}V^{-1}e +\\ &\quad T(e^TV^{-1}e)^{-1}(e^T\mathbf{d}^2V^{-1}e + 2e^T\mathbf{d}V^{-1}\mathbf{d}e)\\ &= \mathbf{tr}\mathbf{d}V^{-1}\mathbf{d}V + \mathbf{tr}V^{-1}\mathbf{d}^2V - T(e^TV^{-1}e)^{-2}(e^T\mathbf{d}V^{-1}e)\mathbf{d}(e^TV^{-1}e) +\\ &\quad T(e^TV^{-1}e)^{-1}e^T\mathbf{d}^2V^{-1}e + 2T(e^TV^{-1}e)^{-1}e^T\mathbf{d}V^{-1}\mathbf{d}e\end{aligned}$$

or

$$\mathbf{d}^2S = \mathbf{tr}\mathbf{d}V^{-1}\mathbf{d}V + \mathbf{tr}V^{-1}\mathbf{d}^2V - \frac{1}{T}\left(\frac{(e^T\mathbf{d}V^{-1}e)}{s^2}\right)^2 + \frac{e^T\mathbf{d}^2V^{-1}e}{s^2} +$$

$$-2\frac{e^T\mathbf{d}V^{-1}X(X^TV^{-1}X)^{-1}X^T\mathbf{d}V^{-1}e}{s^2} \tag{4.6}$$

This expression contains both differentials of V and of V^{-1}. In the sequel we will show how $\mathbf{d}^2S$ can be written as a function of one of these forms.

The inverse of information matrix, defined as $-E\,\dfrac{\partial^2S(\xi)}{\partial\xi\partial\xi^T}$, is of interest because it gives us a lower bound for a consistent estimator of ξ. As Magnus (1978) shows, this expression is equal to minus half of the derivative corresponding to the first part of this expression, $-1/2\mathbf{tr}\mathbf{d}V^{-1}\mathbf{d}V$.

Therefore it has to be positive:

$$\begin{aligned}-\mathbf{tr}\mathbf{d}V^{-1}\mathbf{d}V &= \mathbf{tr}V^{-1}\mathbf{d}VV^{-1}\mathbf{d}V = \mathbf{vec}(V^{-1}\mathbf{d}VV^{-1})\mathbf{vec}(\mathbf{d}V)\\ &= \mathbf{vec}(\mathbf{d}V)^T(V^{-1}\otimes V^{-1})\mathbf{vec}(\mathbf{d}V)\\ &\geq 0.\end{aligned}$$

Furthermore it is obvious that also the third term and the last one, if present, are always non-positive. This is a far from encouraging situation as we are looking for a minimum. On the other hand the signs of the second and fourth term are not clear without any information about the structure of V. We will show that at least in the MA case these expressions are always positive. Anyway, a stationary point need not to be a minimum, nor is a local minimum a global one.

When an expression for V is known, rewrite $\mathbf{d}V^{-1}$ and $\mathbf{d}^2V^{-1}$ as functions of $\mathbf{d}V$ and $\mathbf{d}^2V$:

$\mathbf{d}V^{-1} = -V^{-1}\mathbf{d}VV^{-1}$

$\mathbf{d}^2V^{-1} = \mathbf{d}(\mathbf{d}V^{-1}) = \mathbf{d}(-V^{-1}\mathbf{d}VV^{-1}) = 2V^{-1}\mathbf{d}VV^{-1}\mathbf{d}VV^{-1} - V^{-1}\mathbf{d}^2VV^{-1}$.

Substituting into (4.6) and writing $\phi = V^{-1}e$ we get

$$\mathbf{d}^2S = -\mathbf{tr}V^{-1}\mathbf{d}VV^{-1}\mathbf{d}V + \mathbf{tr}V^{-1}\mathbf{d}^2V - \frac{1}{T}\left(\frac{\phi^T\mathbf{d}V\phi}{s^2}\right)^2 + \frac{\phi^T\mathbf{d}VV^{-1}\mathbf{d}V\phi}{s^2} +$$

$$-\frac{\phi^T\mathbf{d}^2V\phi}{s^2} - 2\frac{\phi^T\mathbf{d}VV^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}\mathbf{d}V\phi}{s^2}$$

or

$$\mathbf{d}^2S = -\mathbf{tr}V^{-1}\mathbf{d}VV^{-1}\mathbf{d}V + \mathbf{tr}V^{-1}\mathbf{d}^2V - \frac{1}{T}\left(\frac{\phi^T\mathbf{d}V\phi}{s^2}\right)^2 - \frac{\phi^T\mathbf{d}^2V\phi}{s^2}$$

$$+ 2\frac{\phi^T\mathbf{d}V(V^{-1} - V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1})\mathbf{d}V\phi}{s^2}. \tag{4.7}$$

Here the last expression at the right hand side is positive, while the signs of the second and fourth one are not clear. On the other hand, when an expression for V^{-1} is known, namely in the pure AR case, it is profitable to rewrite $\mathbf{d}V$ and $\mathbf{d}^2V$: $\mathbf{d}V = -V\mathbf{d}V^{-1}V$ and $\mathbf{d}^2V = 2V\mathbf{d}V^{-1}V\mathbf{d}V^{-1}V - V\mathbf{d}^2V^{-1}V$.

Substitution in (4.6) gives

$$\mathbf{d}^2S = -\mathbf{tr}\mathbf{d}V^{-1}V\mathbf{d}V^{-1}V + \mathbf{tr}V^{-1}(2V\mathbf{d}V^{-1}V\mathbf{d}V^{-1}V - V\mathbf{d}^2V^{-1}V) - \frac{1}{T}\left(\frac{e^T\mathbf{d}V^{-1}e}{s^2}\right)^2 +$$

$$\frac{e^T\mathbf{d}^2V^{-1}e}{s^2} - 2\frac{e^T\mathbf{d}V^{-1}X(X^TV^{-1}X)^{-1}X^T\mathbf{d}V^{-1}e}{s^2}$$

or

$$\mathbf{d}^2S = \mathbf{tr}V\mathbf{d}V^{-1}V\mathbf{d}V^{-1} - \mathbf{tr}V\mathbf{d}^2V^{-1} - \frac{1}{T}\left(\frac{e^T\mathbf{d}V^{-1}e}{s^2}\right)^2 + \frac{e^T\mathbf{d}^2V^{-1}e}{s^2}$$
$$-2\frac{e^T\mathbf{d}V^{-1}X(X^TV^{-1}X)^{-1}X^T\mathbf{d}V^{-1}e}{s^2} \tag{4.8}$$

Observe that the third and last term are always negative. The signs of the second and fourth one are unknown. Next we will use these expressions to give a detailed description of the second derivatives for the ARMA loglikelihood.

4.4 The second derivative

As can be expected the expressions for the second derivatives are quite complicated and the proofs correspondingly tedious. To simplify as much as possible we start by giving some lemmas.

Lemma 4.1

For symmetric A, B $\in \mathbb{R}^{T\times T}$,

$$\mathbf{tr}A\frac{\partial V_{AR}^{-1}}{\partial\vartheta_i}B\frac{\partial V_{AR}^{-1}}{\partial\vartheta_j} =$$
$$\sum_{k=0}^{p}\sum_{l=0}^{p}\sum_{g=1}^{T-k-i}\sum_{h=1}^{T-l-j}\{A[g+k,h+j]B[g+i,h+l] + A[g+k,h+l]B[g+i,h+j] +$$
$$A[g+i,h+j]B[g+k,h+l] + A[g+i,h+l]B[g+k,h+j]\}\vartheta_k\vartheta_l.$$

Proof

Using proposition 3.5 we get

$$\mathbf{tr}A\frac{\partial V_{AR}^{-1}}{\partial\vartheta_i}B\frac{\partial V_{AR}^{-1}}{\partial\vartheta_j} = \mathbf{tr}A\{\sum_{k=0}^{p}(L_{k-i}(k) + L_{k-i}^T(k))\vartheta_k\}B\{\sum_{l=0}^{p}(L_{l-j}(l) + L_{l-j}^T(l))\vartheta_l\}$$
$$= \sum_{k=0}^{p}\sum_{l=0}^{p}\mathbf{tr}\{AL_{k-i}(k)BL_{l-j}(l) + AL_{k-i}(k)BL_{l-j}^T(l) + AL_{k-i}^T(k)BL_{l-j}(l) +$$
$$AL_{k-i}^T(k)BL_{l-j}^T(l)\}\vartheta_k\vartheta_l.$$

For the consecutive parts of the sum we have

$$\mathbf{tr}(AL_{k\text{-}i}(k)BL_{l\text{-}j}(l)) = \mathbf{tr}(A \sum_{g=1+k}^{T\text{-}i} \iota_g \iota_{g\text{-}k+i}^T B \sum_{h=1+l}^{T\text{-}j} \iota_h \iota_{h\text{-}l+j}^T)$$

$$= \sum_{g=1+k}^{T\text{-}i} \sum_{h=1+l}^{T\text{-}j} (\iota_{h\text{-}l+j}^T A \iota_g)(\iota_{g\text{-}k+i}^T B \iota_h)$$

$$= \sum_{g=1+k}^{T\text{-}i} \sum_{h=1+l}^{T\text{-}j} A[h-l+j,g]B[g-k+i,h]$$

$$= \sum_{g=1}^{T\text{-}k\text{-}i} \sum_{h=1}^{T\text{-}l\text{-}j} A[g+k,h+j]B[g+i,h+l],$$

$$\mathbf{tr}AL_{k\text{-}i}(k)BL_{l\text{-}j}^T(l) = \mathbf{tr}A \sum_{g=1+k}^{T\text{-}i} \iota_g \iota_{g\text{-}k+i}^T B \sum_{h=1+l}^{T\text{-}j} (\iota_h \iota_{h\text{-}l+j}^T)^T =$$

$$= \sum_{g=1+k}^{T\text{-}i} \sum_{h=1+l}^{T\text{-}j} \iota_h^T A \iota_g \iota_{g\text{-}k+i}^T B \ \iota_{h\text{-}l+j}$$

$$= \sum_{g=1}^{T\text{-}i\text{-}k} \sum_{h=1}^{T\text{-}j\text{-}l} \iota_{h+l}^T A \iota_{g+k} \iota_{g+i}^T B \ \iota_{h+j}$$

$$= \sum_{g=1}^{T\text{-}i\text{-}k} \sum_{h=1}^{T\text{-}j\text{-}l} A[g+k,h+l]B[g+i,h+j],$$

$$\mathbf{tr}AL_{k\text{-}i}^T(k)BL_{l\text{-}j}(l) = \mathbf{tr}AL_{i\text{-}k}(i)BL_{l\text{-}j}(l) = \sum_{g=1}^{T\text{-}i\text{-}k} \sum_{h=1}^{T\text{-}l\text{-}j} A[g+i,h+j]B[g+k,h+l]$$

and

$$\mathbf{tr}AL_{k\text{-}i}^T(k)BL_{l\text{-}j}^T(l) = \mathbf{tr}AL_{i\text{-}k}(i)BL_{j\text{-}l}(j) = \sum_{g=1}^{T\text{-}i\text{-}k} \sum_{h=1}^{T\text{-}j\text{-}l} A[g+i,h+l]B[g+k,h+j].$$

Together:

$$\mathbf{tr}A\frac{\partial V_{AR}^{-1}}{\partial \vartheta_i}B\frac{\partial V_{AR}^{-1}}{\partial \vartheta_j} = \sum_{k=0}^{p} \sum_{l=0}^{p} \sum_{g=1}^{T\text{-}k\text{-}i} \sum_{h=1}^{T\text{-}l\text{-}j}$$

$$\{A[g+k,h+j]B[g+i,h+l] + A[g+k,h+l]B[g+i,h+j] +$$
$$A[g+i,h+j]B[g+k,h+l] + A[g+i,h+l]B[g+k,h+j]\}\vartheta_k\vartheta_l. \ \square$$

Lemma 4.2

Let $x \in \mathbb{R}^{T\times 1}$ and $A \in \mathbb{R}^{T\times T}$, then

$$x^T\frac{\partial V_{AR}^{-1}}{\partial\vartheta_i}A\frac{\partial V_{AR}^{-1}}{\partial\vartheta_j}x = \sum_{k=0}^{p}\sum_{h=0}^{p}\sum_{s=1}^{T-i-k}\sum_{t=1}^{T-j-h} A[s+i,t+h]x_{s+k}x_{t+j} + A[s+i,t+j]x_{s+k}x_{t+h} + A[s+k,t+h]x_{s+i}x_{t+j} + A[s+k,t+j]x_{s+i}x_{t+h}\vartheta_k\vartheta_h.$$

Proof

Use proposition 3.5 to substitute the derivative

$$x^T\frac{\partial V_{AR}^{-1}}{\partial\vartheta_i}A\frac{\partial V_{AR}^{-1}}{\partial\vartheta_j}x = x^T\{\sum_{k=0}^{p}(L_{k-i}(k)+L_{k-i}^T(k))\vartheta_k\}A\{\sum_{h=0}^{p}(L_{h-j}(h)+L_{h-j}^T(h))\vartheta_h\}x$$

$$= \sum_{k=0}^{p}\sum_{h=0}^{p} x^T\{L_{k-i}(k)+L_{i-k}(i)\}A\{L_{h-j}(h)+L_{j-h}(j)\}x\vartheta_k\vartheta_h$$

$$= \sum_{k=0}^{p}\sum_{h=0}^{p} x^T\{L_{k-i}(k)AL_{h-j}(h)+L_{k-i}(k)AL_{j-h}(j)+L_{i-k}(i)AL_{h-j}(h)+L_{i-k}(i)AL_{j-h}(j)\}x\vartheta_k\vartheta_h.$$

Here $x^TL_{k-i}(k)AL_{h-j}(h)x =$

$$= \sum_{s=1+k}^{T-i}\sum_{t=1+h}^{T-j} x^T\iota_s\iota_{s-k+i}^TA\iota_t\iota_{t-h+j}^Tx$$

$$= \sum_{s=1+k}^{T-i}\sum_{t=1+h}^{T-j} x_sA[s-k+i,t]x_{t-h+j}$$

$$= \sum_{s=1}^{T-i-k}\sum_{t=1}^{T-j-h} A[s+i,t+h]x_{s+k}x_{t+j}.$$

In the same way we get

$$x^TL_{k-i}(k)AL_{j-h}(j)x = \sum_{s=1}^{T-i-k}\sum_{t=1}^{T-h-j} A[s+i,t+j]x_{s+k}x_{t+h},$$

$$x^TL_{i-k}(i)AL_{h-j}(h)x = \sum_{s=1}^{T-k-i}\sum_{t=1}^{T-j-h} A[s+k,t+h]x_{s+i}x_{t+j}$$

and

$$x^T L_{i-k}(i) A L_{j-h}(j) x = \sum_{s=1}^{T-i-k} \sum_{t=1}^{T-h-j} A[s+k,t+j] x_{s+i} x_{t+h}.$$

The result is

$$x^T \frac{\partial V_{AR}^{-1}}{\partial \vartheta_i} A \frac{\partial V_{AR}^{-1}}{\partial \vartheta_j} x = \sum_{k=0}^{p} \sum_{h=0}^{p} \sum_{s=1}^{T-i-k} \sum_{t=1}^{T-j-h} A[s+i,t+h] x_{s+k} x_{t+j} + A[s+i,t+j] x_{s+k} x_{t+h} +$$
$$A[s+k,t+h] x_{s+i} x_{t+j} + A[s+k,t+j] x_{s+i} x_{t+h} \vartheta_k \vartheta_h. \ \square$$

4.4.1 AR-part

In the proof of theorem 4.1 it was shown that we have to start from $V = [N\ M]\bar{V}_{AR}[N\ M]^T$ with $\bar{V}_{AR} = (\bar{P}^T\bar{P} - \bar{Q}\bar{Q}^T)^{-1}$ when we need the differential of $[N\ M]\bar{V}_{AR}\mathbf{d}(\bar{V}_{AR}^{-1})\bar{V}_{AR}[N\ M]^T$. Hence the second differential is a less friendly expression with two terms:
$\mathbf{d}^2V = 2[N\ M]\bar{V}_{AR}\mathbf{d}(\bar{V}_{AR}^{-1})\bar{V}_{AR}\mathbf{d}(\bar{V}_{AR}^{-1})\bar{V}_{AR}[N\ M]^T - [N\ M]\bar{V}_{AR}\mathbf{d}^2(\bar{V}_{AR}^{-1})\bar{V}_{AR}[N\ M]^T$.
It implicates that the resulting second derivatives become correspondingly longer.

Theorem 4.2
Second derivative ARMA-case; AR-part.
Let $\phi = V^{-1}e \in \mathbb{R}^{T\times 1}$,
$U = \bar{V}_{AR}[N\ M]^T V^{-1}[N\ M]\bar{V}_{AR} \in \mathbb{R}^{(T+q)\times(T+q)}$,
$\zeta = \bar{V}_{AR}[N\ M]^T\phi \in \mathbb{R}^{(T+q)\times 1}$,
$H_1 = \bar{V}_{AR}[N\ M]^T(V^{-1} - V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1})[N\ M]\bar{V}_{AR} \in \mathbb{R}^{(T+q)\times(T+q)}$.
Then

$$\frac{\partial^2 S}{\partial\vartheta_i\partial\vartheta_j} = -\mathbf{tr}V^{-1}\frac{\partial V}{\partial\vartheta_i}V^{-1}\frac{\partial V}{\partial\vartheta_j} + \mathbf{tr}V^{-1}\frac{\partial^2 V}{\partial\vartheta_i\partial\vartheta_j} - \frac{1}{T}\frac{1}{s^4}\phi^T\frac{\partial V}{\partial\vartheta_i}\phi\ \phi^T\frac{\partial V}{\partial\vartheta_j}\phi - \frac{2}{s^2}\phi^T\frac{\partial^2 V}{\partial\vartheta_i\partial\vartheta_j}\phi$$
$$+\frac{2}{s^2}\phi^T\frac{\partial V}{\partial\vartheta_i}H\frac{\partial V}{\partial\vartheta_j}\phi \tag{4.9}$$

with

$$\mathbf{tr}V^{-1}\frac{\partial V}{\partial\vartheta_i}V^{-1}\frac{\partial V}{\partial\vartheta_j} = 2\sum_{k=0}^{p}\sum_{l=0}^{p}\sum_{g=1}^{T+q-i-k}\sum_{h=1}^{T+q-j-l}$$
$$(U[g+k,h+j]U[g+i,h+l] + U[g+k,h+l]U[g+i,h+j])\vartheta_k\vartheta_l \tag{4.9.1}$$

$$\mathbf{tr}V^{-1}\frac{\partial^2 V}{\partial\vartheta_i\partial\vartheta_j} = 2\mathbf{tr}U\frac{\partial\bar{V}_{AR}^{-1}}{\partial\vartheta_i}\bar{V}_{AR}\frac{\partial\bar{V}_{AR}^{-1}}{\partial\vartheta_j} - \mathbf{tr}U\frac{\partial^2\bar{V}_{AR}^{-1}}{\partial\vartheta_i\partial\vartheta_j} \tag{4.9.2}$$

$$\mathbf{tr}U\frac{\partial\bar{V}_{AR}^{-1}}{\partial\vartheta_i}\bar{V}_{AR}\frac{\partial\bar{V}_{AR}^{-1}}{\partial\vartheta_j} = \sum_{k=0}^{p}\sum_{h=0}^{p}\sum_{g=1}^{T+q-k-i}\sum_{h=1}^{T+q-l-j}$$
$$\{U[g+k,h+j]\bar{V}_{AR}[g-h+i-l] + U[g+k,h+l]\bar{V}_{AR}[g-h+i-j] + U[g+i,h+j]\bar{V}_{AR}[g-h+k-l] + U[g+i,h+l]\bar{V}_{AR}[g-h+k-j]\}\vartheta_k\vartheta_l. \tag{4.9.2a}$$

$$\mathbf{tr}U\frac{\partial^2\bar{V}_{AR}^{-1}}{\partial\vartheta_i\partial\vartheta_j} = 2\sum_{h=1}^{T+q-i-j} U[h+i,h+j] \tag{4.9.2b}$$

$$\phi^T\frac{\partial\bar{V}_{AR}^{-1}}{\partial\vartheta_i}\phi\ \phi^T\frac{\partial\bar{V}_{AR}^{-1}}{\partial\vartheta_j}\phi = 4\{\sum_{k=0}^{p}\sum_{g=1}^{T+q-i-k}\zeta_{g+k}\zeta_{g+i}\vartheta_k\}\{\sum_{l=0}^{p}\sum_{h=1}^{T+q-j-l}\zeta_{h+l}\zeta_{h+j}\vartheta_l\} \tag{4.9.3}$$

$$\phi^T\frac{\partial^2 V}{\partial\vartheta_i\partial\vartheta_j}\phi = 2\zeta^T\frac{\partial\bar{V}_{AR}^{-1}}{\partial\vartheta_i}\bar{V}_{AR}\frac{\partial\bar{V}_{AR}^{-1}}{\partial\vartheta_j}\zeta - \zeta^T\frac{\partial^2\bar{V}_{AR}^{-1}}{\partial\vartheta_i\partial\vartheta_j}\zeta \tag{4.9.4}$$

$$\zeta^T\frac{\partial\bar{V}_{AR}^{-1}}{\partial\vartheta_i}\bar{V}_{AR}\frac{\partial\bar{V}_{AR}^{-1}}{\partial\vartheta_j}\zeta = \sum_{k=0}^{p}\sum_{h=0}^{p}\sum_{s=1}^{T+q-i-k}\sum_{t=1}^{T+q-j-h}$$
$$\bar{V}_{AR}[s-t+i-h]\zeta_{s+k}^T\zeta_{t+j} + \bar{V}_{AR}[s-t+i-j]\zeta_{s+k}^T\zeta_{t+h} + \bar{V}_{AR}[s-t+k-h]\zeta_{s+i}^T\zeta_{t+j} + \bar{V}_{AR}[s-t+k-j]\zeta_{s+i}^T\zeta_{t+h}\vartheta_k\vartheta_h \tag{4.9.4a}$$

$$\zeta^T\frac{\partial^2\bar{V}_{AR}^{-1}}{\partial\vartheta_i\partial\vartheta_j}\zeta = \sum_{h=1}^{T+q-i-j}\zeta_{h+j}\zeta_{h+i} \tag{4.9.4b}$$

$$\phi^T\frac{\partial V}{\partial\vartheta_i}H\frac{\partial V}{\partial\vartheta_j}\phi = \sum_{k=0}^{p}\sum_{h=0}^{p}\sum_{s=1}^{T+q-i-k}\sum_{t=1}^{T+q-j-h} H_1[s+i,t+h]\zeta_{s+k}^T\zeta_{t+j} +$$
$$H_1[s+i,t+j]\zeta_{s+k}^T\zeta_{t+h} + H_1[s+k,t+h]\zeta_{s+i}^T\zeta_{t+j} + H_1[s+k,t+j]\zeta_{s+i}^T\zeta_{t+h}\vartheta_k\vartheta_h. \tag{4.9.5}$$

Proof

From proposition 3.5 we have the differential of the inverse of the AR covariance matrix, $\mathbf{d}\bar{V}_{AR}^{-1}$. For the information matrix, $-1/2\mathbf{tr}V^{-1}\mathbf{d}VV^{-1}\mathbf{d}V$, we get, substituting $\mathbf{d}V = -[N\ M]\bar{V}_{AR}\mathbf{d}\bar{V}_{AR}^{-1}\bar{V}_{AR}[N\ M]^T$:

$$\begin{aligned}\mathbf{tr}V^{-1}\mathbf{d}VV^{-1}\mathbf{d}V &= \mathbf{tr}V^{-1}\ [N\ M]\bar{V}_{AR}\mathbf{d}\bar{V}_{AR}^{-1}\bar{V}_{AR}[N\ M]^T\ V^{-1}\ [N\ M]\bar{V}_{AR}\mathbf{d}\bar{V}_{AR}^{-1}\bar{V}_{AR}[N\ M]^T\\ &= \mathbf{tr}\bar{V}_{AR}[N\ M]^TV^{-1}[N\ M]\bar{V}_{AR}\ \mathbf{d}\bar{V}_{AR}^{-1}\ \bar{V}_{AR}[N\ M]^TV^{-1}[N\ M]\bar{V}_{AR}\ \mathbf{d}\bar{V}_{AR}^{-1}\\ &= \mathbf{tr}U\mathbf{d}\bar{V}_{AR}^{-1}U\mathbf{d}\bar{V}_{AR}^{-1}.\end{aligned}$$

Because of lemma 4.1 we get, replacing T by T+q and substituting A=B=U:

$$\mathbf{tr}U\frac{\partial\bar{V}_{AR}^{-1}}{\partial\vartheta_i}U\frac{\partial\bar{V}_{AR}^{-1}}{\partial\vartheta_j} =$$

$$2\sum_{k=0}^{p}\sum_{l=0}^{p}\sum_{g=1}^{T+q-i-k}\sum_{h=1}^{T+q-j-l}(U[g+k,h+j]U[g+i,h+l]+U[g+k,h+l]U[g+i,h+j])\vartheta_k\vartheta_l.$$

For the second part of the second differential, first substitute $\mathbf{d}^2V$:

$$\begin{aligned}\mathbf{tr}V^{-1}\mathbf{d}^2V &= 2\mathbf{tr}V^{-1}[N\ M]\bar{V}_{AR}\mathbf{d}\bar{V}_{AR}^{-1}\bar{V}_{AR}\mathbf{d}\bar{V}_{AR}^{-1}\bar{V}_{AR}[N\ M]^T - \mathbf{tr}V^{-1}[N\ M]\bar{V}_{AR}\mathbf{d}^2\bar{V}_{AR}^{-1}\bar{V}_{AR}[N\ M]^T\\ &= 2\mathbf{tr}\bar{V}_{AR}[N\ M]^TV^{-1}[N\ M]\bar{V}_{AR}\ \mathbf{d}\bar{V}_{AR}^{-1}\ \bar{V}_{AR}\ \mathbf{d}\bar{V}_{AR}^{-1} - \mathbf{tr}\bar{V}_{AR}[N\ M]^TV^{-1}[N\ M]\bar{V}_{AR}\ \mathbf{d}^2\bar{V}_{AR}^{-1}\\ &= 2\mathbf{tr}U\mathbf{d}\bar{V}_{AR}^{-1}\bar{V}_{AR}\mathbf{d}\bar{V}_{AR}^{-1}\ - \mathbf{tr}U\mathbf{d}^2\bar{V}_{AR}^{-1}.\end{aligned}$$

The first component of the second part is, using lemma 4.1 with T+q instead of T and A=U and $B=\bar{V}_{AR}$:

$$\mathbf{tr}U\frac{\partial\bar{V}_{AR}^{-1}}{\partial\vartheta_i}\bar{V}_{AR}\frac{\partial\bar{V}_{AR}^{-1}}{\partial\vartheta_j} = \sum_{k=0}^{p}\sum_{h=0}^{p}\sum_{g=1}^{T+q-k-i}\sum_{h=1}^{T+q-l-j}\{U[g+k,h+j]\bar{V}_{AR}[g-h+i-l]+U[g+k,h+l]$$
$$\bar{V}_{AR}[g-h+i-j]+U[g+i,h+j]\bar{V}_{AR}[g-h+k-l]+U[g+i,h+l]\bar{V}_{AR}[g-h+k-j]\}\vartheta_k\vartheta_l.$$

The second component of the second part is less complicated. Here we have

$$\mathbf{tr}U\frac{\partial^2\bar{V}_{AR}^{-1}}{\partial\vartheta_i\partial\vartheta_j} = \mathbf{tr}\{U\{(L_{j-i}(j)+L_{j-i}^T(j)\}\} = 2\mathbf{tr}\{UL_{j-i}(j)\}$$

$$= 2\mathbf{tr}\{U\sum_{h=1+j}^{T+q-i}\iota_h\iota_{h-j+i}^T) = 2\sum_{h=1+j}^{T+q-i}U[h-j+i,h] = 2\sum_{h=1}^{T+q-i-j}U[h+i,h+j],$$

the elements of the $|i-j|^{th}$ diagonal of U.

For the quadratic part we need $\phi^T\mathbf{d}V\phi$ or $-\phi^T[N\ M]\bar{V}_{AR}\mathbf{d}\bar{V}_{AR}^{-1}\bar{V}_{AR}[N\ M]^T\phi$. Therefore the quadratic form becomes:

$$\zeta^T \frac{\partial \bar{V}_{AR}^{-1}}{\partial \vartheta_i} \zeta = \zeta^T \{ \sum_{k=0}^{p} (L_{k-i}(k) + L_{k-i}^T(k)) \vartheta_k \} \zeta$$

$$= 2 \sum_{k=0}^{p} \zeta^T \sum_{g=1+k}^{T+q-i} \iota_g \iota_{g-k+i}^T \zeta \ \vartheta_k$$

$$= 2 \sum_{k=0}^{p} \sum_{g=1}^{T+q-i-k} \zeta_{g+k} \zeta_{g+i} \vartheta_k.$$

The third part of the second derivative is, apart of a constant,

$$\zeta^T \frac{\partial \bar{V}_{AR}^{-1}}{\partial \vartheta_i} \zeta \ \zeta^T \frac{\partial \bar{V}_{AR}^{-1}}{\partial \vartheta_j} \zeta = 4 \{ \sum_{k=0}^{p} \sum_{g=1}^{T+q-i-k} \zeta_{g+k} \zeta_{g+i} \vartheta_k \} \{ \sum_{l=0}^{p} \sum_{h=1}^{T+q-j-l} \zeta_{h+l} \zeta_{h+j} \vartheta_l \}.$$

The fourth part, $\phi^T \mathbf{d}^2 V \phi$, has again two components:

$$\phi^T \mathbf{d}^2 V \phi = 2\phi^T [N\ M] \bar{V}_{AR} \mathbf{d} \bar{V}_{AR}^{-1} \bar{V}_{AR} \mathbf{d} \bar{V}_{AR}^{-1} \bar{V}_{AR} [N\ M]^T \phi - \phi^T [N\ M] \bar{V}_{AR} \mathbf{d}^2 \bar{V}_{AR}^{-1} \bar{V}_{AR} [N\ M]^T \phi$$
$$= 2\zeta^T \mathbf{d} \bar{V}_{AR}^{-1} \bar{V}_{AR} \mathbf{d} \bar{V}_{AR}^{-1} \zeta \ - \zeta^T \mathbf{d}^2 \bar{V}_{AR}^{-1} \zeta.$$

For the former of these $\zeta^T \mathbf{d} \bar{V}_{AR}^{-1} \bar{V}_{AR} \mathbf{d} \bar{V}_{AR}^{-1} \zeta$, we get because of lemma 4.2 with $T+q$ instead of T, and $A = \bar{V}_{AR}$:

$$\zeta^T \frac{\partial \bar{V}_{AR}^{-1}}{\partial \vartheta_i} \bar{V}_{AR} \frac{\partial \bar{V}_{AR}^{-1}}{\partial \vartheta_j} \zeta = \sum_{k=0}^{p} \sum_{h=0}^{p} \sum_{s=1}^{T+q-i-k} \sum_{t=1}^{T+q-j-h} \bar{V}_{AR}[s-t+i-h] \zeta_{s+k} \zeta_{t+j} +$$
$$\bar{V}_{AR}[s-t+i-j] \zeta_{s+k} \zeta_{t+h} + \bar{V}_{AR}[s-t+k-h] \zeta_{s+i} \zeta_{t+j} + \bar{V}_{AR}[s-t+k-j] \zeta_{s+i} \zeta_{t+h} \vartheta_k \vartheta_h$$

For the latter part of the fourth part we find a simple expression:

$$\zeta^T \frac{\partial^2 \bar{V}_{AR}^{-1}}{\partial \vartheta_i \partial \vartheta_j} \zeta = \zeta^T \{ L_{j-i}(j) + L_{j-i}^T(j) \} \zeta = 2\zeta^T \{ L_{j-i}(j) \} \zeta =$$

$$= 2\zeta^T \sum_{h=1+j}^{T+q-i} \iota_h \iota_{h-j+i}^T \zeta = \sum_{h=1+j}^{T+q-i} \zeta_h \zeta_{h-j+i} = \sum_{h=1}^{T+q-i-j} \zeta_{h+j} \zeta_{h+i}.$$

For the last part substitute $\mathbf{d}V$ and use $\zeta = \bar{V}_{AR} [N\ M]^T \phi$, as defined above:

$$\phi^T \mathbf{d}V (V^{-1} - V^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1}) \mathbf{d}V \phi =$$
$$= \phi^T ([N\ M] \bar{V}_{AR} \mathbf{d} \bar{V}_{AR}^{-1} \bar{V}_{AR} [N\ M]^T) (V^{-1} - V^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1})$$
$$([N\ M] \bar{V}_{AR} \mathbf{d} \bar{V}_{AR}^{-1} \bar{V}_{AR} [N\ M]^T) \phi$$
$$= \zeta^T \mathbf{d} \bar{V}_{AR}^{-1} \bar{V}_{AR} [N\ M]^T (V^{-1} - V^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1}) [N\ M] \bar{V}_{AR} \mathbf{d} \bar{V}_{AR}^{-1} \zeta$$

$= \zeta^T \mathbf{d}\bar{V}_{AR}^{-1} H_1 \mathbf{d}\bar{V}_{AR}^{-1}\zeta.$

In view of lemma 4.2:

$$\phi^T \frac{\partial V}{\partial \vartheta_i} H \frac{\partial V}{\partial \vartheta_j}\phi = \sum_{k=0}^{p}\sum_{h=0}^{p}\sum_{s=1}^{T+q-i-k}\sum_{t=1}^{T+q-j-h} H_1[s+i,t+h]\zeta_{s+k}^T\zeta_{t+j} + H_1[s+i,t+j]\zeta_{s+k}^T\zeta_{t+h} + H_1[s+k,t+h]\zeta_{s+i}^T\zeta_{t+j} + H_1[s+k,t+j]\zeta_{s+i}^T\zeta_{t+h}\vartheta_k\vartheta_h. \ \square$$

4.4.2 MA-part

From theorem (4.7) and proposition (3.3) we find the second derivatives of the modified loglikelihood to α.

Theorem 4.3

Second derivative ARMA-case; MA-part.

Let $\phi = V^{-1}e \in \mathbb{R}^{T\times 1}$ and

$H_2 = (V^{-1} - V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}) \in \mathbb{R}^{T\times T}$.

Then

$$\frac{\partial^2 S}{\partial\alpha_i\partial\alpha_j} = -\mathbf{tr}V^{-1}\frac{\partial V}{\partial\alpha_i}V^{-1}\frac{\partial V}{\partial\alpha_j} + \mathbf{tr}V^{-1}\frac{\partial^2 V}{\partial\alpha_i\partial\alpha_j} - \frac{1}{T}\frac{1}{s^4}\phi^T\frac{\partial V}{\partial\alpha_i}\phi\ \phi^T\frac{\partial V}{\partial\alpha_j}\phi + -\frac{2}{s^2}\phi^T\frac{\partial^2 V}{\partial\alpha_i\partial\alpha_j}\phi + \frac{2}{s^2}\phi^T\frac{\partial V}{\partial\alpha_i}H_2\frac{\partial V}{\partial\alpha_j}\phi \tag{4.10}$$

with

$$\mathbf{tr}V^{-1}\frac{\partial V}{\partial\alpha_i}V^{-1}\frac{\partial V}{\partial\alpha_j} = 2\sum_{k=0}^{q}\sum_{l=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T}\sum_{s=1}^{T}\sum_{t=1}^{T} \bar{V}_{AR}[h-g-i+k]\bar{V}_{AR}[s-t-i+l](V^{-1}[t,h]V^{-1}[g,s] + V^{-1}[s,h]V^{-1}[g,t])\alpha_k\alpha_l \tag{4.10.1}$$

$$\mathbf{tr}V^{-1}\frac{\partial^2 V}{\partial\alpha_i\partial\alpha_j} = 2\sum_{h=1}^{T}\sum_{k=1}^{T}\bar{V}_{AR}[h-k-i+j]V^{-1}[k,h] \tag{4.10.2}$$

$$\phi^T\frac{\partial V}{\partial\alpha_i}\phi\phi^T\frac{\partial V}{\partial\alpha_j}\phi = 4\left(\sum_{k=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T}\bar{V}_{AR}[h-g-i+k]\phi_g\phi_h\alpha_k\right)$$
$$\left(\sum_{k=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T}\bar{V}_{AR}[h-g-j+k]\phi_g\phi_h\alpha_k\right) \qquad (4.10.3)$$

$$\phi^T\frac{\partial^2 V}{\partial\alpha_i\partial\alpha_j}\phi = 2\sum_{h=1}^{T}\sum_{g=1}^{T}\bar{V}_{AR}[h-g-i+j]\phi_h\phi_g \qquad (4.10.4)$$

$$\phi^T\frac{\partial V}{\partial\alpha_i}H_2\frac{\partial V}{\partial\alpha_j}\phi = \sum_{h=1}^{T}\sum_{g=1}^{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\left(\sum_{k=0}^{q}\sum_{l=0}^{q}\bar{V}_{AR}[h-g-i+k]\bar{V}_{AR}[s-t-j+l]\alpha_k\alpha_l\right)$$
$$(H_2[g,s]\phi_h\phi_t + H_2[g,t]\phi_h\phi_s + H_2[h,s]\phi_g\phi_t + H_2[h,t]\phi_g\phi_s), \qquad (4.10.5)$$

Proof

To find the second derivative substitute in (4.3) the components of the second derivatives. Again, the first part is the most complicated and computer time consuming one.

$$\mathbf{tr}V^{-1}\frac{\partial V}{\partial\alpha_i}V^{-1}\frac{\partial V}{\partial\alpha_j} =$$

$$= \mathbf{tr}V^{-1}\left(\sum_{k=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T}\bar{V}_{AR}[h-g-i+k](\iota_h\iota_g^T+\iota_g\iota_h^T)\alpha_k\right)V^{-1}$$
$$\left(\sum_{l=0}^{q}\sum_{s=1}^{T}\sum_{t=1}^{T}\bar{V}_{AR}[s-t-i+l](\iota_s\iota_t^T+\iota_t\iota_s^T)\alpha_l\right)$$
$$= \sum_{k=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T}\sum_{l=0}^{q}\sum_{s=1}^{T}\sum_{t=1}^{T}\bar{V}_{AR}[h-g-i+k]\bar{V}_{AR}[s-t-i+l]$$
$$\mathbf{tr}(V^{-1}\iota_h\iota_g^T+V^{-1}\iota_g\iota_h^T)(V^{-1}\iota_s\iota_t^T+V^{-1}\iota_t\iota_s^T)\alpha_k\alpha_l$$
$$= 2\sum_{k=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T}\sum_{l=0}^{q}\sum_{s=1}^{T}\sum_{t=1}^{T}\bar{V}_{AR}[h-g-i+k]\bar{V}_{AR}[s-t-i+l]$$
$$(V^{-1}[t,h]V^{-1}[g,s]+V^{-1}[s,h]V^{-1}[g,t])\alpha_k\alpha_l,$$

because

$$\mathbf{tr}(V^{-1}\iota_h\iota_g^T+V^{-1}\iota_g\iota_h^T)(V^{-1}\iota_s\iota_t^T+V^{-1}\iota_t\iota_s^T) =$$
$$= \mathbf{tr}(V^{-1}\iota_h\iota_g^TV^{-1}\iota_s\iota_t^T+V^{-1}\iota_h\iota_g^TV^{-1}\iota_t\iota_s^T+V^{-1}\iota_g\iota_h^TV^{-1}\iota_s\iota_t^T+V^{-1}\iota_g\iota_h^TV^{-1}\iota_t\iota_s^T)$$

$$= V^{-1}[t,h]V^{-1}[g,s] + V^{-1}[s,h]V^{-1}[g,t] + V^{-1}[t,g]V^{-1}[h,s] + V^{-1}[s,g]V^{-1}[h,t]$$
$$= 2(V^{-1}[t,h]V^{-1}[g,s] + V^{-1}[s,h]V^{-1}[g,t]).$$

The next part is easy and consists of the elements of $\bar{V}_{AR}$ and V^{-1}:

$$\mathbf{tr}V^{-1}\frac{\partial^2 V}{\partial\alpha_i\partial\alpha_j} = \mathbf{tr}V^{-1}\sum_{h=1}^{T}\sum_{k=1}^{T}\bar{V}_{AR}[h-k-i+j](\iota_h\iota_k^T + \iota_k\iota_h^T) =$$
$$= \sum_{h=1}^{T}\sum_{k=1}^{T}\bar{V}_{AR}[h-k-i+j]\mathbf{tr}V^{-1}(\iota_h\iota_k^T + \iota_k\iota_h^T)$$
$$= \sum_{h=1}^{T}\sum_{k=1}^{T}\bar{V}_{AR}[h-k-i+j]\iota_k^T V^{-1}\iota_h + \iota_h^T V^{-1}\iota_k)$$
$$= 2\sum_{h=1}^{T}\sum_{k=1}^{T}\bar{V}_{AR}[h-k-i+j]V^{-1}[k,h].$$

For the quadratic part we first compute $\phi^T\frac{\partial V}{\partial\alpha_i}\phi$, with $\phi = V^{-1}e$.

$$\phi^T\frac{\partial V}{\partial\alpha_i}\phi = \phi^T(\sum_{j=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T}\bar{V}_{AR}[h-g-i+j](\iota_h\iota_g^T + \iota_g\iota_h^T)\alpha_j)\phi$$
$$= \sum_{j=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T}\bar{V}_{AR}[h-g-i+j](\phi^T\iota_h\iota_g^T\phi + \phi^T\iota_g\iota_h^T\phi)\alpha_j$$
$$= 2\sum_{j=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T}\bar{V}_{AR}[h-g-i+j]\phi_g\phi_h\alpha_j.$$

The third component becomes

$$\phi^T\frac{\partial V}{\partial\alpha_i}\phi\phi^T\frac{\partial V}{\partial\alpha_j}\phi = 4(\sum_{k=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T}\bar{V}_{AR}[h-g-i+k]\phi_g\phi_h\alpha_k)$$
$$(\sum_{k=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T}\bar{V}_{AR}[h-g-j+k]\phi_g\phi_h\alpha_k).$$

The quadratic part with the second derivative of V is the weighted sum of the elements of ϕ:

$$\phi^T \frac{\partial^2 V}{\partial\alpha_i \partial\alpha_j}\phi = 2\sum_{h=1}^{T}\sum_{g=1}^{T} \bar{V}_{AR}[h-g-i+j]\phi_h\phi_g.$$

For the last part we define H_2 as before and get

$$\phi^T \frac{\partial V}{\partial\alpha_i} H_2 \frac{\partial V}{\partial\alpha_j}\phi = \phi^T \sum_{k=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T} \bar{V}_{AR}[h\text{-}g\text{-}i+k](\iota_h\iota_g^T + \iota_g\iota_h^T)\alpha_k \; H_2$$

$$\sum_{l=0}^{q}\sum_{s=1}^{T}\sum_{t=1}^{T} \bar{V}_{AR}[s\text{-}t\text{-}j+l](\iota_s\iota_t^T + \iota_t\iota_s^T)\alpha_l\phi$$

$$= \sum_{k=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T}\sum_{l=0}^{q}\sum_{s=1}^{T}\sum_{t=1}^{T} \bar{V}_{AR}[h\text{-}g\text{-}i+k]\bar{V}_{AR}[s\text{-}t\text{-}j+l]\phi^T(\iota_h\iota_g^T + \iota_g\iota_h^T)\; H_2$$
$$(\iota_s\iota_t^T + \iota_t\iota_s^T)\phi\alpha_k\alpha_l$$

$$= \sum_{h=1}^{T}\sum_{g=1}^{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\Big(\sum_{k=0}^{q}\sum_{l=0}^{q} \bar{V}_{AR}[h\text{-}g\text{-}i+k]\bar{V}_{AR}[s\text{-}t\text{-}j+l]\alpha_k\alpha_l\Big)$$
$$(H_2[g,s]\phi_h\phi_t + H_2[g,t]\phi_h\phi_s + H_2[h,s]\phi_g\phi_t + H_2[h,t]\phi_g\phi_s),$$

because $(\phi^T\iota_h\iota_g^T + \phi^T\iota_g\iota_h^T)\; H_2\; (\iota_s\iota_t^T\phi + \iota_t\iota_s^T\phi) =$

$$= \phi^T\iota_h\iota_g^T H_2\iota_s\iota_t^T\phi + \phi^T\iota_h\iota_g^T H_2\iota_t\iota_s^T\phi + \phi^T\iota_g\iota_h^T H_2\iota_s\iota_t^T\phi + \phi^T\iota_g\iota_h^T H_2\iota_t\iota_s^T\phi$$
$$= H_2[g,s]\phi_h\phi_t + H_2[g,t]\phi_h\phi_s + H_2[h,s]\phi_g\phi_t + H_2[h,t]\phi_g\phi_s. \quad \square$$

4.4.3 AR/MA-part

Theorem 4.4

Second derivative ARMA-case: AR/MA-part

Let $\phi = V^{-1}e \in \mathbb{R}^{T\times 1}$,

$Z = \bar{V}_{AR}[N\ M]^T V^{-1} \in \mathbb{R}^{(T+q)\times T}$,

$\zeta = Ze \in \mathbb{R}^{(T+q)\times 1}$, and

$H_3 = (V^{-1} - V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1})[N\ M]\bar{V}_{AR} \in \mathbb{R}^{T\times(T+q)}$.

Then

$$\frac{\partial^2 S}{\partial\alpha_i\partial\vartheta_j} = -\mathbf{tr}Z^T\frac{\partial\bar{V}_{AR}^{-1}}{\partial\vartheta_i}Z\frac{\partial V}{\partial\alpha_j} + \mathbf{tr}V^{-1}\frac{\partial^2 V}{\partial\alpha_i\partial\vartheta_j} - \frac{1}{T}\frac{1}{s^4}\phi^T\frac{\partial V}{\partial\alpha_i}\phi\; \zeta^T\frac{\partial\bar{V}_{AR}^{-1}}{\partial\vartheta_j}\zeta +$$

$$-\frac{1}{s^2}\phi^T\frac{\partial^2 V}{\partial\alpha_i\partial\vartheta_j}\phi + \phi^T\frac{\partial V}{\partial\alpha_i}H_3\frac{\partial\bar{V}_{AR}^{-1}}{\partial\vartheta_j}\zeta, \tag{4.11}$$

where

$$\mathbf{tr}Z^T\frac{\partial \bar{V}_{AR}^{-1}}{\partial \vartheta_i}Z\frac{\partial V}{\partial \alpha_j} = 2\sum_{k=0}^{p}\sum_{l=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T}\sum_{t=1}^{T+q-k-i}\bar{V}_{AR}[h-g-i+l]$$
$$(Z[t+k,g]Z[t+i,h]+Z[t+k,h]Z[t+i,g])\vartheta_k\alpha_l \tag{4.11.1}$$

$$\mathbf{tr}V^{-1}\frac{\partial^2 V}{\partial \alpha_i \partial \vartheta_j} = -2\sum_{l=0}^{p}\sum_{k=0}^{p}\sum_{h=1}^{T}\sum_{g=1}^{T}\sum_{t=1}^{T+q-k-l}$$
$$(\bar{V}_{AR}[h-i-t-k]\bar{V}_{AR}[g-j-t-l]+\bar{V}_{AR}[h-i-t-l]\bar{V}_{AR}[g-j-t-k])V^{-1}[h,g]\vartheta_k\alpha_j \tag{4.11.2}$$

$$\phi^T\frac{\partial V}{\partial \alpha_i}\phi \ \zeta^T\frac{\partial \bar{V}_{AR}^{-1}}{\partial \vartheta_j}\zeta = 4\Big(\sum_{j=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T}\bar{V}_{AR}[h-g-i+j]\phi_g\phi_h\alpha_j\Big)\Big(\sum_{k=0}^{p}\sum_{g=1}^{T+q-i-k}\zeta_{g+k}\zeta_{g+i}\vartheta_k\Big) \tag{4.11.3}$$

$$\phi^T\frac{\partial^2 V}{\partial \alpha_i \partial \vartheta_j}\phi = -2\sum_{l=0}^{p}\sum_{k=0}^{p}\sum_{h=1}^{T}\sum_{g=1}^{T}\sum_{t=1}^{T+q-k-l}$$
$$(\bar{V}_{AR}[h-i-t-k]\bar{V}_{AR}[g-j-t-l]+\bar{V}_{AR}[h-i-t-l]\bar{V}_{AR}[g-j-t-k])\phi_h^T\phi_g\vartheta_k\alpha_j \tag{4.11.4}$$

$$\phi^T\frac{\partial V}{\partial \alpha_i}H_3\frac{\partial \bar{V}_{AR}^{-1}}{\partial \vartheta_j}\zeta = \sum_{l=0}^{q}\sum_{k=0}^{p}\sum_{s=1}^{T}\sum_{t=1}^{T}\Big(\sum_{g=1}^{T+q-j-k}\bar{V}_{AR}[s-t-i+l](\phi_t H_3[s,g+k]\zeta_{g+j} +$$
$$\phi_t H_3[s,g+j]\zeta_{g+k} + \phi_s H_3[t,g+k]\zeta_{g+j} + \phi_s H_3[t,g+j])\zeta_{g+k}\alpha_l\vartheta_k. \tag{4.11.5}$$

Proof

Here we write $\mathbf{d}_\alpha$ and $\mathbf{d}_\vartheta$ to make clear which differential is meant. From proposition 3.3 we conclude

$$\mathbf{d}_\vartheta\mathbf{d}_\alpha V = \sum_{i=1}^{q}\sum_{j=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T}\mathbf{d}\bar{V}_{AR}[h-g-i+j](\iota_h\iota_g^T+\iota_g\iota_h^T)\alpha_j\mathbf{d}\alpha_i,$$

where

$$\mathbf{d}\bar{V}_{AR}[h-g-i+j] = \mathbf{d}\iota_{h-i}^T\bar{V}_{AR}\iota_{g-j} = -\iota_{h-i}^T\bar{V}_{AR}\mathbf{d}_\vartheta\bar{V}_{AR}^{-1}\bar{V}_{AR}\iota_{g-j} =$$

$$= -\iota_{h-i}^T\bar{V}_{AR}\sum_{l=0}^{p}\sum_{k=0}^{p}(L_{k-l}(k)+L_{k-l}^T(k))\vartheta_k\mathbf{d}\vartheta_l\bar{V}_{AR}\iota_{g-j}$$

$$= -\sum_{l=0}^{p}\sum_{k=0}^{p}\iota_{h-i}^T\bar{V}_{AR}\Big(\sum_{t=1+k}^{T+q-l}\bar{\iota}_t\bar{\iota}_{t-k+l}^T+\bar{\iota}_{t-k+l}\bar{\iota}_t^T\Big)\bar{V}_{AR}\iota_{g-j}\vartheta_k\mathbf{d}\vartheta_l$$

$$= \sum_{l=0}^{p} \sum_{k=0}^{p} \sum_{t=1}^{T+q-k-l} (\bar{V}_{AR}[h\text{-}i\text{-}t\text{-}k]\bar{V}_{AR}[g\text{-}j\text{-}t\text{-}l] + \bar{V}_{AR}[h\text{-}i\text{-}t\text{-}l]\bar{V}_{AR}[g\text{-}j\text{-}t\text{-}k])\vartheta_k \mathbf{d}\vartheta_l.$$

The result is

$$\mathbf{d}_{\vartheta}\mathbf{d}_{\alpha}V = -\sum_{i=1}^{q} \sum_{j=0}^{q} \sum_{l=0}^{p} \sum_{k=0}^{p} \sum_{h=1}^{T} \sum_{g=1}^{T} \sum_{t=1}^{T+q-k-l}$$
$$(\bar{V}_{AR}[h\text{-}i\text{-}t\text{-}k]\bar{V}_{AR}[g\text{-}j\text{-}t\text{-}l] + \bar{V}_{AR}[h\text{-}i\text{-}t\text{-}l]\bar{V}_{AR}[g\text{-}j\text{-}t\text{-}k])(\iota_h\iota_g^T + \iota_g\iota_h^T)\vartheta_k\alpha_j\mathbf{d}\vartheta_l\mathbf{d}\alpha_i$$

For the first part of the second differential of the modified likelihood function we have

$$\mathbf{tr}\mathbf{d}_{\vartheta}V^{-1}\mathbf{d}_{\alpha}V = -\mathbf{tr}V^{-1}\mathbf{d}V_{\vartheta}V^{-1}\mathbf{d}_{\alpha}V$$
$$= \mathbf{tr}V^{-1}[N\ M]\bar{V}_{AR}\mathbf{d}_{\vartheta}\bar{V}_{AR}^{-1}\bar{V}_{AR}[N\ M]^TV^{-1}\mathbf{d}_{\alpha}V$$
$$= \mathbf{tr}Z^T\mathbf{d}_{\vartheta}\bar{V}_{AR}^{-1}Z\mathbf{d}_{\alpha}V$$

with $Z = \bar{V}_{AR}[N\ M]^TV^{-1} \in \mathbb{R}^{(T+q)xT}$.

The corresponding derivative, using propositions 3.3 and 3.5, is

$$\mathbf{tr}Z^T\frac{\partial\bar{V}_{AR}^{-1}}{\partial\vartheta_i}Z\frac{\partial V}{\partial\alpha_j} =$$

$$\mathbf{tr}\{Z^T\sum_{k=0}^{p}(L_{k\text{-}i}(k) + L_{k\text{-}i}^T(k))\vartheta_k\}Z\sum_{l=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T}\bar{V}_{AR}[h\text{-}g\text{-}i+l](\iota_h\iota_g^T + \iota_g\iota_h^T)\alpha_l$$

$$= \mathbf{tr}\sum_{k=0}^{p}\sum_{l=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T}\bar{V}_{AR}[h\text{-}g\text{-}i+l](Z^TL_{k\text{-}i}(k)Z + Z^TL_{k\text{-}i}^T(k)Z)(\iota_h\iota_g^T + \iota_g\iota_h^T)\vartheta_k\alpha_l$$

$$= 2\mathbf{tr}\sum_{k=0}^{p}\sum_{l=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T}\sum_{t=1}^{T+q-k-i}\bar{V}_{AR}[h\text{-}g\text{-}i+l](Z^T\bar{\iota}_{t+k}\bar{\iota}_{t+i}^TZ)(\iota_h\iota_g^T + \iota_g\iota_h^T)\vartheta_k\alpha_l$$

$$= 2\sum_{k=0}^{p}\sum_{l=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T}\sum_{t=1}^{T+q-k-i}$$
$$\bar{V}_{AR}[h\text{-}g\text{-}i+l](Z[t+k,g]Z[t+i,h] + Z[t+k,h]Z[t+i,g])\vartheta_k\alpha_l,$$

because $\bar{V}_{AR}[.]$ is a scalar and

$$\mathbf{tr}(Z^T\bar{\iota}_{t+k}\bar{\iota}_{t+i}^TZ)(\iota_h\iota_g^T + \iota_g\iota_h^T) =$$
$$= \iota_g^TZ^T\bar{\iota}_{t+k}\bar{\iota}_{t+i}^TZ\iota_h + \iota_h^TZ^T\bar{\iota}_{t+k}\bar{\iota}_{t+i}^TZ\iota_g$$
$$= (\bar{\iota}_{t+k}^TZ\iota_g)^T\ \bar{\iota}_{t+i}^TZ\iota_h + (\bar{\iota}_{t+k}^TZ\iota_h)^T\ \bar{\iota}_{t+i}^TZ\iota_g$$

$= Z[t+k,g]Z[t+i,h] + Z[t+k,h]Z[t+i,g].$

For the second term, $\mathbf{tr}V^{-1}\frac{\partial^2 V}{\partial\alpha_i\partial\vartheta_j}$, we get

$$\mathbf{tr}V^{-1}\frac{\partial^2 V}{\partial\alpha_i\partial\vartheta_j} = -\mathbf{tr}V^{-1}\sum_{l=0}^{p}\sum_{k=0}^{p}\sum_{h=1}^{T}\sum_{g=1}^{T}\sum_{t=1}^{T+q-k-l}$$
$$(\bar{V}_{AR}[h-i-t-k]\bar{V}_{AR}[g-j-t-l] + \bar{V}_{AR}[h-i-t-l]\bar{V}_{AR}[g-j-t-k])(\iota_h\iota_g^T + \iota_g\iota_h^T)\vartheta_k\alpha_j$$
$$= -2\sum_{l=0}^{p}\sum_{k=0}^{p}\sum_{h=1}^{T}\sum_{g=1}^{T}\sum_{t=1}^{T+q-k-l}$$
$$(\bar{V}_{AR}[h-i-t-k]\bar{V}_{AR}[g-j-t-l] + \bar{V}_{AR}[h-i-t-l]\bar{V}_{AR}[g-j-t-k])V^{-1}[h,g]\vartheta_k\alpha_j,$$

as $\mathbf{tr}V^{-1}(\iota_h\iota_g^T + \iota_g\iota_h^T) = V^{-1}[g,h] + V^{-1}[h,g] = 2V^{-1}[h,g]$.

For the third term we need $\phi^T\mathbf{d}_\vartheta V\phi$ and $\phi^T\mathbf{d}_\alpha V\phi$. As before we use $\phi = V^{-1}e$ and $\zeta = \bar{V}_{AR}[N\ M]^T\phi$. For the former one we have

$$\phi^T\frac{\partial V}{\partial\vartheta_j}\phi = -\zeta^T\frac{\partial\bar{V}_{AR}^{-1}}{\partial\vartheta_j}\zeta = 2\sum_{k=0}^{p}\sum_{g=1}^{T+q-i-k}\zeta_{g+k}\zeta_{g+i}\vartheta_k,$$

for the latter one

$$\phi^T\frac{\partial V}{\partial\alpha_i}\phi = 2\sum_{j=0}^{q}\sum_{h=1}^{T}\sum_{g=1}^{T}\bar{V}_{AR}[h-g-i+j]\phi_g\phi_h\alpha_j.$$

For the fourth term we conclude from the second differential and $\phi^T(\iota_h\iota_g^T + \iota_g\iota_h^T)\phi = 2\phi_h^T\phi_g$:

$$\phi^T\frac{\partial^2 V}{\partial\alpha_i\partial\vartheta_j}\phi = -2\sum_{l=0}^{p}\sum_{k=0}^{p}\sum_{h=1}^{T}\sum_{g=1}^{T}\sum_{t=1}^{T+q-k-l}(\bar{V}_{AR}[h-i-t-k]\bar{V}_{AR}[g-j-t-l] +$$
$$\bar{V}_{AR}[h-i-t-l]\bar{V}_{AR}[g-j-t-k])\phi_h^T\phi_g\vartheta_k\alpha_j.$$

The last term is the derivative corresponding to
$\phi^T\mathbf{d}V(V^{-1} - V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1})\mathbf{d}V\phi$
or more precisely
$\phi^T\mathbf{d}_\alpha V(V^{-1} - V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1})[N\ M]\bar{V}_{AR}\mathbf{d}_\vartheta\bar{V}_{AR}^{-1}\bar{V}_{AR}[N\ M]^T\phi.$
For the derivatives we have

$$\frac{\partial V}{\partial \alpha_i}\phi = \sum_{l=0}^{q} \sum_{s=1}^{T} \sum_{t=1}^{T} \bar{V}_{AR}[s\text{-}t\text{-}i+l](\iota_s \iota_t^T + \iota_t \iota_s^T)\alpha_l \ \phi$$

$$= \sum_{l=0}^{q} \sum_{s=1}^{T} \sum_{t=1}^{T} \bar{V}_{AR}[s\text{-}t\text{-}i+l](\phi_t \iota_s + \phi_s \iota_t)\alpha_l$$

and

$$\frac{\partial \bar{V}_{AR}^{-1}}{\partial \vartheta_j}\zeta = \sum_{k=0}^{p} (L_{k\text{-}j}(k) + L_{k\text{-}j}^T(k))\vartheta_k \zeta$$

$$= \sum_{k=0}^{p} \{ \sum_{g=1+k}^{T+q\text{-}j} \bar{\iota}_g \bar{\iota}_{g\text{-}k+j}^T + (\bar{\iota}_g \bar{\iota}_{g\text{-}k+j}^T)^T \vartheta_k \}\zeta$$

$$= \sum_{k=0}^{p} (\sum_{g=1}^{T+q\text{-}j\text{-}k} \vartheta_k (\zeta_{g+j} \bar{\iota}_{g+k} + \zeta_{g+k} \bar{\iota}_{g+j}).$$

The result is

$$\phi^T \frac{\partial V}{\partial \alpha_i} H_3 \frac{\partial \bar{V}_{AR}^{-1}}{\partial \vartheta_j}\zeta = \sum_{l=0}^{q} \sum_{s=1}^{T} \sum_{t=1}^{T} \bar{V}_{AR}[s\text{-}t\text{-}i+l](\phi_t \iota_s^T + \phi_s \iota_t^T)\alpha_l \cdot H_3 \cdot$$

$$\sum_{k=0}^{p} (\sum_{g=1}^{T+q\text{-}j\text{-}k} \vartheta_k (\zeta_{g+j} \bar{\iota}_{g+k} + \zeta_{g+k} \bar{\iota}_{g+j})$$

$$= \sum_{l=0}^{q} \sum_{k=0}^{p} \sum_{s=1}^{T} \sum_{t=1}^{T} (\sum_{g=1}^{T+q\text{-}j\text{-}k} \bar{V}_{AR}[s\text{-}t\text{-}i+l](\phi_t \iota_s^T + \phi_s \iota_t^T) H_3 (\zeta_{g+j} \bar{\iota}_{g+k} + \zeta_{g+k} \bar{\iota}_{g+j})\alpha_l \vartheta_k$$

$$= \sum_{l=0}^{q} \sum_{k=0}^{p} \sum_{s=1}^{T} \sum_{t=1}^{T} (\sum_{g=1}^{T+q\text{-}j\text{-}k} \bar{V}_{AR}[s\text{-}t\text{-}i+l] A \alpha_l \vartheta_k,$$

where

$$A = (\phi_t \iota_s^T + \phi_s \iota_t^T) H_3 (\zeta_{g+j} \bar{\iota}_{g+k} + \zeta_{g+k} \bar{\iota}_{g+j}) =$$
$$= \phi_t H_3[s,g+k]\zeta_{g+j} \ + \phi_t H_3[s,g+j]\zeta_{g+k} \ + \phi_s H_3[t,g+k]\zeta_{g+j} \ + \phi_s H_3[t,g+j]\zeta_{g+k}.$$

$$\phi^T \frac{\partial V}{\partial \alpha_i} H_3 \frac{\partial \bar{V}_{AR}^{-1}}{\partial \vartheta_j}\zeta = \sum_{l=0}^{q} \sum_{k=0}^{p} \sum_{s=1}^{T} \sum_{t=1}^{T} (\sum_{g=1}^{T+q\text{-}j\text{-}k} \bar{V}_{AR}[s\text{-}t\text{-}i+l](\phi_t H_3[s,g+k]\zeta_{g+j}$$
$$+ \phi_t H_3[s,g+j]\zeta_{g+k} \ + \phi_s H_3[t,g+k]\zeta_{g+j} \ + \phi_s H_3[t,g+j]\zeta_{g+k}\zeta_{g+k}\alpha_l \vartheta_k). \ \square$$

4.4.4 Pure AR case

For the second derivative of the pure AR case we start from (4.8) as an expression for V^{-1} in stead of V is now available. From the end of section 2.4 we know that in the pure AR case the determinant of the covariance matrix is equal to its (p×p) left upper submatrix. Comparing the differential of both - proposition (3.5) and proposition (3.7) - we conclude that the submatrix has a much more complicated form. Hence it is more profitable to use the expression for the covariance matrix and to replace T by 2p, as the value of the determinant of the covariance matrix is independent of its dimensions.

Corollary 4.3

Second derivative AR-case.

$$\frac{\partial^2 S}{\partial\vartheta_i\partial\vartheta_j} = \mathbf{tr}V\frac{\partial V^{-1}}{\partial\vartheta_i}V\frac{\partial V^{-1}}{\partial\vartheta_j} - \mathbf{tr}V\frac{\partial^2 V^{-1}}{\partial\vartheta_i\partial\vartheta_j} - \frac{1}{T}\frac{1}{s^4}\, e^T\frac{\partial V^{-1}}{\partial\vartheta_i}e\; e^T\frac{\partial V^{-1}}{\partial\vartheta_j}e \; +$$

$$+\frac{1}{s^2}\, e^T\frac{\partial^2 V^{-1}}{\partial\vartheta_i\partial\vartheta_j}e - \frac{2}{s^2}\, e^T\frac{\partial V^{-1}}{\partial\vartheta_i}\{X(X^TV^{-1}X)^{-1}X^T\}\frac{\partial V^{-1}}{\partial\vartheta_j}e, \tag{4.12}$$

where

$$\mathbf{tr}V\frac{\partial V^{-1}}{\partial\vartheta_i}V\frac{\partial V^{-1}}{\partial\vartheta_j} = \sum_{k=0}^{p}\sum_{l=0}^{p}\sum_{g=1}^{2p-k-i}\sum_{h=1}^{2p-l-j}$$
$$\{V[g-h-j+k]V[g-h+i-l]+V[g-h+k-l]V[g-h+i-j]\}\vartheta_k\vartheta_l \tag{4.12.1}$$

$$\mathbf{tr}V\frac{\partial^2 \bar{V}_{AR}^{-1}}{\partial\vartheta_i\partial\vartheta_j} = 2\sum_{h=1}^{2p-i-j} V[i-j] = 2(2p-i-j)\ V[i-j] \tag{4.12.2}$$

$$e^T\frac{\partial V^{-1}}{\partial\vartheta_i}e\; e^T\frac{\partial V^{-1}}{\partial\vartheta_j}e = 4\{\sum_{k=0}^{p}(\sum_{h=1}^{T-i-k} e_{h+k}e_{h+i})\vartheta_k\}\{\sum_{k=0}^{p}(\sum_{h=1}^{T-j-k} e_{h+k}e_{h+j})\vartheta_k\} \tag{4.12.3}$$

$$e^T\frac{\partial^2 V^{-1}}{\partial\vartheta_i\partial\vartheta_j}e = 2\sum_{h=1}^{T-i-j} e_{h+j}e_{h+i} \tag{4.12.4}$$

$$\phi^T\frac{\partial V}{\partial\vartheta_i}H\frac{\partial V}{\partial\vartheta_j}\phi = \sum_{k=0}^{p}\sum_{h=0}^{p}\sum_{s=1}^{T-i-k}\sum_{t=1}^{T-j-h} H_4[s+i,t+h]e_{s+k}^Te_{t+j}+H_4[s+i,t+j]e_{s+k}^Te_{t+h}+$$
$$H_4[s+k,t+h]e_{s+i}^Te_{t+j}+H_4[s+k,t+j]e_{s+i}^Te_{t+h}\vartheta_k\vartheta_h. \qquad (4.12.5)$$

Proof

Starting from lemma 4.1 and using $A=B=V$, $T=2p$ and $q=0$ we get expression (4.12.1) and from (4.9.2b) we obtain, as $U=V_{AR}V_{AR}^{-1}V_{AR}=V_{AR}$ expression (4.12.2).

For expression (4.12.3) we use (4.9.3) and the fact that $\zeta=\bar{V}_{AR}\phi=e$; expression (4.12.4) follows directly from (4.9.4b).

Eventually (4.12.5) results from lemma 4.2 with $A=X(X^TV^{-1}X)^{-1}X^T=H_4$.

Remark

For the determinant part we also can use $\underline{V}$, the p×p left upper part of V. From the differential

$$\mathbf{d}\underline{V}^{-1} = \sum_{i=0}^{p}\Big(\sum_{j=0}^{p-i-1}\{L_{j-i}(j)+L_{j-i}^T(j)\}\vartheta_j - \sum_{j=p-i+1}^{p}\{L_{j-i}(p-i)+L_{j-i}^T(p-i)\}\vartheta_j\Big)\mathbf{d}\vartheta_i,$$

we get the derivatives

$$\frac{\partial\underline{V}^{-1}}{\partial\vartheta_i} = \sum_{j=0}^{p-i-1}\{L_{j-i}(j)+L_{j-i}^T(j)\}\vartheta_j - \sum_{j=p-i+1}^{p}\{L_{j-i}(p-i)+L_{j-i}^T(p-i)\}\vartheta_j),\ i=1,..p$$

and

$$\frac{\partial^2\underline{V}^{-1}}{\partial\vartheta_i\partial\vartheta_j} = \begin{cases} L_{j-i}(j)+L_{j-i}^T(j) & 0\le j\le p-i-1 \\ L_{j-i}(p-i)+L_{j-i}^T(p-i) & p-i+1\le j\le p. \end{cases}$$

The first term of the second derivative of the AR modified likelihood function is rather complicated. The derivation takes several steps. First, use the derivative of the submatrix.

$$\mathbf{tr}\underline{V}\frac{\partial\underline{V}^{-1}}{\partial\vartheta_i}\underline{V}\frac{\partial\underline{V}^{-1}}{\partial\vartheta_j} =$$

$$\mathbf{tr}\underline{V}\Big\{\sum_{k=0}^{p-i-1}\{L_{k-i}(k)+L_{k-i}^T(k)\}\vartheta_k - \sum_{k=p-i+1}^{p}\{L_{k-i}(p-i)+L_{k-i}^T(p-i)\}\vartheta_k\Big\}\times$$

$$\underline{V}\{\sum_{l=0}^{p-j-1}\{L_{l-j}(l)+L^{T}_{l-j}(l)\}\vartheta_l - \sum_{l=p-j+1}^{p}\{L_{l-j}(p-j)+L^{T}_{l-j}(p-j)\}\vartheta_l\}$$

Multiplication gives:

$$\mathbf{tr}\underline{V}\frac{\partial\underline{V}^{-1}}{\partial\vartheta_i}\underline{V}\frac{\partial\underline{V}^{-1}}{\partial\vartheta_j} = \mathbf{tr}\{\sum_{k=0}^{p-i-1}\sum_{l=0}^{p-j-1}(\underline{V}L_{k-i}(k)+\underline{V}L^{T}_{k-i}(k))\ (\underline{V}L_{l-j}(l)+\underline{V}L^{T}_{l-j}(l))\vartheta_k\vartheta_l +$$

$$-\sum_{k=0}^{p-i-1}\sum_{l=p-j+1}^{p}(\underline{V}L_{k-i}(k)+\underline{V}L^{T}_{k-i}(k))\ (\underline{V}L_{l-j}(p-j)+\underline{V}L^{T}_{l-j}(p-j))\vartheta_k\vartheta_l +$$

$$-\sum_{k=p-i+1}^{p}\sum_{l=0}^{p-j-1}(\underline{V}L_{k-i}(p-i)+\underline{V}L^{T}_{k-i}(p-i))\ (\underline{V}L_{l-j}(l)+\underline{V}L^{T}_{l-j}(l))\vartheta_k\vartheta_l +$$

$$+\sum_{k=p-i+1}^{p}\sum_{l=p-j+1}^{p}(\underline{V}L_{k-i}(p-i)+\underline{V}L^{T}_{k-i}(p-i))\ (\underline{V}L_{l-j}(p-j)+\underline{V}L^{T}_{l-j}(p-j))\vartheta_k\vartheta_l\}.$$

As all terms contain similar expressions we compute next

$$\begin{aligned}
&\mathbf{tr}\{\underline{V}L_{k_1}(k_2)+\underline{V}L^{T}_{k_1}(k_2)\}\{\underline{V}L_{l_1}(l_2)+\underline{V}L^{T}_{l_1}(l_2)\} = \\
&= \mathbf{tr}\{\underline{V}L_{k_1}(k_2)\underline{V}L_{l_1}(l_2)+\underline{V}L_{k_1}(k_2)\underline{V}L^{T}_{l_1}(l_2)+\underline{V}L^{T}_{k_1}(k_2)\underline{V}L_{l_1}(l_2)+\underline{V}L^{T}_{k_1}(k_2)\underline{V}L^{T}_{l_1}(l_2)\} \\
&= 2\mathbf{tr}\{\underline{V}L_{k_1}(k_2)\underline{V}L_{l_1}(l_2)+\underline{V}L_{k_1}(k_2)\underline{V}L^{T}_{l_1}(l_2)\} \\
&= 2\sum_{g=1+k_2}^{p+k_1-k_2}\sum_{h=1+l_2}^{p+l_1-l_2}\mathbf{tr}\{\underline{V}\iota_g\iota^{T}_{g-k_1}\underline{V}\iota_h\iota^{T}_{h-l_1}+\underline{V}\iota_g\iota^{T}_{g-k_1}\underline{V}(\iota_h\iota^{T}_{h-l_1})^{T}\} \\
&= 2\sum_{g=1+k_2}^{p+k_1-k_2}\sum_{h=1+l_2}^{p+l_1-l_2}\iota^{T}_{h-l_1}\underline{V}\iota_g\iota^{T}_{g-k_1}\underline{V}\iota_h+\iota^{T}_{h}\underline{V}\iota_g\iota^{T}_{g-k_1}\underline{V}\iota_{h-l_1} \\
&= 2\sum_{g=1+k_2}^{p+k_1-k_2}\sum_{h=1+l_2}^{p+l_1-l_2}\underline{V}[h-l_1-g]\underline{V}[g-k_1-h]+\underline{V}[h-g]\ \underline{V}[g-k_1-h+l_1].
\end{aligned}$$

Substitute this result:

$$\mathbf{tr}\underline{V}\frac{\partial\underline{V}^{-1}}{\partial\vartheta_i}\underline{V}\frac{\partial\underline{V}^{-1}}{\partial\vartheta_j} = 2\sum_{k=0}^{p-i-1}\sum_{l=0}^{p-j-1}\sum_{g=1+k}^{p-i}\sum_{h=1+l}^{p-j}\{\underline{V}[h-l+j-g]\underline{V}[g-k+i-h]\ +$$

$$\underline{V}[h-g]\ \underline{V}[g-k+i-h+l-j]\}\vartheta_k\vartheta_l\ +$$

$$-2\sum_{k=0}^{p-i-1}\sum_{l=p-j+1}^{p}\sum_{g=1+k}^{p-i}\sum_{h=1+p-j}^{l}\{\underline{V}[h-l+j-g]\underline{V}[g-k+i-h]+$$
$$\underline{V}[h-g]\ \underline{V}[g-k+i-h+l-j]\}\vartheta_k\vartheta_l\ +$$
$$-2\sum_{k=p-i+1}^{p}\sum_{l=0}^{p-j-1}\sum_{g=1+p-i}^{k}\sum_{h=1+l}^{p-j}\{\underline{V}[h-l+j-g]\underline{V}[g-k+i-h]+$$
$$\underline{V}[h-g]\ \underline{V}[g-k+i-h+l-j]\}\vartheta_k\vartheta_l\ +$$
$$+2\sum_{k=p-i+1}^{p}\sum_{l=p-j+1}^{p}\sum_{g=1+p-i}^{k}\sum_{h=1+p-j}^{l}\{\underline{V}[h-l+j-g]\underline{V}[g-k+i-h]+$$
$$\underline{V}[h-g]\ \underline{V}[g-k+i-h+l-j]\}\vartheta_k\vartheta_l$$

and observe that the expression between braces is equal in all four terms. Hence

$$\mathbf{tr}\underline{V}\frac{\partial\underline{V}^{-1}}{\partial\vartheta_i}\underline{V}\frac{\partial\underline{V}^{-1}}{\partial\vartheta_j}=2\{\sum_{k=0}^{p-i-1}\sum_{l=0}^{p-j-1}\sum_{g=1+k}^{p-i}\sum_{h=1+l}^{p-j}-\sum_{k=0}^{p-i-1}\sum_{l=p-j+1}^{p}\sum_{g=1+k}^{p-i}\sum_{h=1+p-j}^{l}$$
$$-\sum_{k=p-i+1}^{p}\sum_{l=0}^{p-j-1}\sum_{g=1+p-i}^{k}\sum_{h=1+l}^{p-j}+\sum_{k=p-i+1}^{p}\sum_{l=p-j+1}^{p}\sum_{g=1+p-i}^{k}\sum_{h=1+p-j}^{l}\}$$
$$\{\underline{V}[h-l+j-g]\underline{V}[g-k+i-h]\ +\underline{V}[h-g]\ \underline{V}[g-k+i-h+l-j]\}\vartheta_k\vartheta_l.$$

The second term becomes

$$\mathbf{tr}\underline{V}\ \frac{\partial^2\underline{V}^{-1}}{\partial\vartheta_i\partial\vartheta_j}=\begin{cases}\mathbf{tr}\underline{V}\ \{L_{j-i}(j)+L^T_{j-i}(j)\} & 0\le j\le p-i-1\\ \mathbf{tr}\underline{V}\ \{L_{j-i}(p-i)+L^T_{j-i}(p-i)\} & p-i+1\le j\le p\end{cases}$$

or as $\underline{V}$ is symmetric:

$$\mathbf{tr}\underline{V}\ \frac{\partial^2\underline{V}^{-1}}{\partial\vartheta_i\partial\vartheta_j}=$$
$$\begin{cases}2\mathbf{tr}\underline{V}\ L_{j-i}(j)\ =\ 2\mathbf{tr}\underline{V}\ \sum_{h=1+j}^{p-i}\iota_h\iota^T_{h-j+i}=2(p-i-j)\underline{V}[j-i] & 0\le j\le p-i-1\\ 2\mathbf{tr}\underline{V}\ L_{j-i}(p-i)=\ 2\mathbf{tr}\underline{V}\ \sum_{h=1+p-i}^{j}\iota_h\iota^T_{h-j+i}=2(j+i-p)\underline{V}[j-i] & p-i+1\le j\le p\end{cases}$$
$$=\ 2|p-i-j|\underline{V}[j-i].\ \square$$

4.4.5 Pure MA-case

Next we will treat the second derivative of the pure MA-case. Before we do so we proof

Lemma 4.3

If $x \in \mathbb{R}^{T\times 1}$ and h-i=g-j, then $\sum_{h=1}^{T}\sum_{g=1}^{T} x_g x_h = \sum_{k=1}^{T-|i-j|} x_{k+|i-j|}x_k.$

Proof

Put $l=h-i=g-j$, then $\sum_{h=1}^{T}\sum_{g=1}^{T} x_g x_h = \sum_{l=\max(1-i,1-j)}^{\min(T-i,T-j)} x_{l+j}x_{l+i}.$

The index l runs from $\max(1-i,1-j)$ to $\min(T-i,T-j)$, or $1-\min(i,j)$ to $T-\max(i,j)$. Substitute $k=l+\min(i,j)$, which runs from 1 to $T-\max(i,j)+\min(i,j)$; the upper bound is equal to $T-i+j$ if $i\geq j$ and $T-j+i$ if $i\leq j$, or $T-|i-j|$. For the indices of x we get $l+j=k-\min(i,j)+j=k+\max(j-i,0)$ and $l+i=k+\max(i-j,0)$, or $x_{k+j-i}x_k$ if $i\leq j$ and $x_k x_{k+i-j}$ if $i>j$. □

Corollary 4.4

Second derivative MA-case.

Let $\phi = V^{-1}e \in \mathbb{R}^{T\times 1}$ and

$H_2 = (V^{-1} - V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}) \in \mathbb{R}^{T\times T}$.

Then

$$\frac{\partial^2 S}{\partial\alpha_i\partial\alpha_j} = -\mathbf{tr}V^{-1}\frac{\partial V}{\partial\alpha_i}V^{-1}\frac{\partial V}{\partial\alpha_j} + \mathbf{tr}V^{-1}\frac{\partial^2 V}{\partial\alpha_i\partial\alpha_j} - \frac{1}{T}\frac{1}{s^4}\phi^T\frac{\partial V}{\partial\alpha_i}\phi\,\phi^T\frac{\partial V}{\partial\alpha_j}\phi - \frac{2}{s^2}\phi^T\frac{\partial^2 V}{\partial\alpha_i\partial\alpha_j}\phi$$
$$+\frac{2}{s^2}\phi^T\frac{\partial V}{\partial\alpha_i}H_2\frac{\partial V}{\partial\alpha_j}\phi \tag{4.13}$$

with

$$\mathbf{tr}V^{-1}\frac{\partial V}{\partial\alpha_i}V^{-1}\frac{\partial V}{\partial\alpha_j} = 2\sum_{k=0}^{q}\sum_{l=0}^{q}\sum_{n=1}^{T-|i-k|}\sum_{m=1}^{T-|i-l|}$$
$$(V^{-1}[m,n+|i-k|)]V^{-1}[m+|i-l|,n] + V^{-1}[m+|i-l|,n]V^{-1}[m,n+|k-i|)a_k a_l \tag{4.13.1}$$

$$\mathbf{tr}V^{-1}\frac{\partial^2 V}{\partial\alpha_i\partial\alpha_j} = 2\sum_{k=1}^{T-|i-j|} V^{-1}[k+|i-j|] \tag{4.13.2}$$

$$\phi^T\frac{\partial V}{\partial\alpha_i}\phi\phi^T\frac{\partial V}{\partial\alpha_j}\phi = 4\Big(\sum_{k=0}^{q}\sum_{n=1}^{T-|i-k|}\phi_{n+|i-k|}\phi_n\alpha_k\Big)\Big(\sum_{k=0}^{q}\sum_{n=1}^{T-|j-k|}\phi_{n+|j-k|}\phi_n\alpha_k\alpha_k\Big) \tag{4.13.3}$$

$$\phi^T\frac{\partial^2 V}{\partial\alpha_i\partial\alpha_j}\phi = 2\sum_{k=1}^{T-|i-j|}\phi_{k+|i-j|}\phi_k \tag{4.13.4}$$

$$\phi^T\frac{\partial V}{\partial\alpha_i}H_2\frac{\partial V}{\partial\alpha_j}\phi = 4\sum_{k=0}^{q}\sum_{l=0}^{q}\sum_{n=1}^{T-|i-k|}\sum_{m=1}^{T-|j-l|}\{\phi_{n+|i-k|}H_2[n,m+|j-l|]\phi_m) + \phi_{n+|i-k|}H_2$$
$$[n,m]\phi_{m+|j-l|} + \phi_nH_2[n+|i-k|,m+|j-l|]\phi_m + \phi_nH_2[n+|i-k|,m]\phi_{m+|j-l|}\}\alpha_k\alpha_l. \tag{4.13.5}$$

Proof

To prove (4.13.1) use (4.10.1) where $\bar{V}_{AR}[h-g-i+k] = \bar{V}_{AR}[s-t-i+l] = I$, which implies $n=h-i=g-k$ and $m=s-i=t-l$. Hence

$$\mathbf{tr}V^{-1}\frac{\partial V}{\partial\alpha_i}V^{-1}\frac{\partial V}{\partial\alpha_j} = 2\sum_{k=0}^{q}\sum_{l=0}^{q}\sum_{n=\max(1-i,1-k)}^{\min(T-i,T-k)}\sum_{m=1+\max(1-i,1-l)}^{\min(T-i,T-l)}$$
$$(V^{-1}[m+l,n+i]V^{-1}[n+k,m+i] + V^{-1}[m+i,n+i]V^{-1}[n+k,m+l])a_ka_l.$$

Next change the indices n and m :

$$\mathbf{tr}V^{-1}\frac{\partial V}{\partial\alpha_i}V^{-1}\frac{\partial V}{\partial\alpha_j} = 2\sum_{k=0}^{q}\sum_{l=0}^{q}\sum_{n=1}^{T-|i-k|}\sum_{m=1}^{T-|i-l|}$$
$$(V^{-1}[m+\max(l-i,0),n+\max(i-k,0)]V^{-1}[m+\max(i-l,0),n+\max(k-i,0)] +$$
$$V^{-1}[m+\max(i-l,0),n+\max(i-k,0)]V^{-1}[m+\max(l-i,0),n+\max(k-i,0)])a_ka_l.$$
$$= 2\sum_{k=0}^{q}\sum_{l=0}^{q}\sum_{n=1}^{T-|i-k|}\sum_{m=1}^{T-|i-l|}$$
$$(V^{-1}[m,n)]V^{-1}[m+|i-l|,n+|i-k|] + V^{-1}[m+|i-l|,n]V^{-1}[m,n+|i-k|)a_ka_l.$$

The last line is a consequence of the fact that in all cases the arguments of V^{-1} are similar, as is shown in the next table.

	i>l	i<l
i>k	(m,n+i-k), (m+i-l,n)	(m+l-i,n+i-k), (m,n)
	(m+i-l,n+i-k), (m,n)	(m,n+i-k), (m+l-i,n)
i<k	(m,n), (m+i-l,n+k-i)	(m+l-i,n), (m,n+k-i)
	(m+i-l,n), (m,n+k-i)	(m,n), (m+l-i,n+k-i)

Expressions (4.13.2), (4.13.3), (4.13.4) follow directly from (4.10.2), (4.10.3), (4.10.4) applying lemma 4.3.

From (4.10.5) we have, putting $n=h\text{-}i=g\text{-}k$ and $m=s\text{-}j=t\text{-}l$:

$$\phi^T \frac{\partial V}{\partial \alpha_i} H_2 \frac{\partial V}{\partial \alpha_j}\phi = \phi^T \sum_{k=0}^{q} \sum_{n=\max(1-i,1-k)}^{\min(T-i,T-k)} (\iota_{n+i}\iota_{n+k}^T + \iota_{n+k}\iota_{n+i}^T)\alpha_k \, H_2$$

$$\sum_{l=0}^{q} \sum_{m=\max(1-j,1-l)}^{\min(T-j,T-l)} (\iota_{m+j}\iota_{m+l}^T + \iota_{m+l}\iota_{m+j}^T)\alpha_l\phi$$

$$= \phi^T \sum_{k=0}^{q} \sum_{n=1}^{T-|i-k|} (\iota_{n+\max(i-k,0)}\iota_{n+\max(k-i,0)}^T + \iota_{n+\max(k-i,0)}\iota_{n+\max(i-k,0)}^T)\alpha_k \, H_2$$

$$\sum_{l=0}^{q} \sum_{m=1}^{T-|j-l|} (\iota_{m+\max(j-l,0)}\iota_{m+\max(l-j,0)}^T + \iota_{m+\max(l-j,0)}\iota_{m+\max(j-l,0)}^T)\alpha_l\phi$$

$$= 4\phi^T \sum_{k=0}^{q} \sum_{n=1}^{T-|i-k|} \sum_{l=0}^{q} \sum_{m=1}^{T-|j-l|}$$

$$(\iota_{n+|i-k|}\iota_n^T + \iota_n\iota_{n+|i-k|}^T)\alpha_k \; H_2 \; (\iota_{m+|j-l|}\iota_m^T + \iota_m\iota_{m+|j-l|}^T)\alpha_l\phi$$

$$= 4\phi^T \sum_{k=0}^{q} \sum_{l=0}^{q} \sum_{n=1}^{T-|i-k|} \sum_{m=1}^{T-|j-l|}$$

$$(\iota_{n+|i-k|}\iota_n^T \, H_2\iota_{m+|j-l|}\iota_m^T \; + \iota_{n+|i-k|}\iota_n^T \, H_2\iota_m\iota_{m+|j-l|}^T \; +$$
$$\iota_n\iota_{n+|i-k|}^T H_2\iota_{m+|j-l|}\iota_m^T \; + \iota_n\iota_{n+|i-k|}^T H_2\iota_m\iota_{m+|j-l|}^T)\alpha_k\alpha_l\phi$$

$$= 4 \sum_{k=0}^{q} \sum_{l=0}^{q} \sum_{n=1}^{T-|i-k|} \sum_{m=1}^{T-|j-l|}$$

$$\{\phi_{n+|i-k|}H_2[n,m+|j-l|]\phi_m) + \phi_{n+|i-k|}H_2[n,m]\phi_{m+|j-l|} +$$
$$\phi_n H_2[n+|i-k|,m+|j-l|]\phi_m + \phi_n H_2[n+|i-k|,m]\phi_{m+|j-l|}\}\alpha_k\alpha_l. \;\square$$

4.5 Conclusion

In this chapter first and second derivatives of the modified likelihood function are derived. Special cases, namely the pure AR and the pure MA case, are treated separately in corollaries, as they are much simpler than the general case. The first derivatives look like linear functions of the ARMA parameters: *e.g.* $\frac{\partial S}{\partial \vartheta} = G\vartheta + g$ for the AR parameters. However the elements of G, for which we give an expression, strongly depend on the parameters of interest. Consequently, a solution of the first order conditions in the form of $\vartheta = G^{-1}g$ is useless. On the other hand the elements of G and g are simple to program using the results of the preceding chapters: they are a function the covariance matrix and some parts of it depending on the specific case. These remarks hold not only for the simple AR and MA cases, but also for the ARMA case.
Furthermore we presented in this chapter expressions for the second derivative of the modified likelihood function and the information matrix. All (five) parts of which the second derivative consists, have a rather simple structure and can easily be programmed. The information matrix is a sum of the elements of a matrix which can be constructed from the covariance matrix.

V ASYMPTOTIC RESULTS

5.1 Approximations for quadratic form and determinant

The concentrated likelihood function (1.12) consists of two parts, a quadratic form and a function of the determinant of the ARMA covariance matrix. Of course both components are as complicated as the covariance matrix itself. One can wonder whether they can be replaced by more simple expressions when the number of observations grows larger. In many textbooks a simpler approach is presented. It amounts to disregarding the starting values $\check{\varepsilon}$ and $\check{v}$, which reduces problems remarkably. The main reason for this procedure may have been the fact that no closed form for the exact covariance matrix was available. We will see that this approach is nevertheless justified if the number of observations is large.
Before we give some theorems regarding the limits of both components, we present a lemma about the inverse of a lower band matrix and a lemma for the limiting behaviour of the quadratic part of the likelihood function. We will show that the limit of the determinant in the likelihood function is a bounded function of T and that it is positive. This implies that its value to the power T^{-1} will tend to one for $T \to \infty$. This gives a justification to drop the determinant part in the likelihood function if T becomes large.

5.2 Some lemmas

First we show how the lower band matrix $M \in \mathbb{R}^{T \times T}$, as defined in section 2.2, containing the MA parameters and its inverse can be written in the form of block Toeplitz matrices. The result will be used in the next lemma 5.2.

Lemma 5.1

Let the lower band matrix $M=B_{T,T}(1,\alpha_1,..,\alpha_q,0,..0)$, $\underline{M}=B_{q,q}(1,\alpha_1,..,\alpha_{q-1})$ and $\underline{N}=B^T_{q,q}(\alpha_q,..,\alpha_1)$. The inverse of M, apart of the last rows and columns, can be written in block form where the $(i,j)^{th}$ block,

$$M^{-1}_{i,j}=\begin{cases}\underline{M}^{-1} & \text{if } i=j\\ (-\underline{M}^{-1}\underline{N})^{i-j}\underline{M}^{-1} & \text{if } i\geq j\\ 0 & \text{if } i<j.\end{cases} \tag{5.1}$$

The eigenvalues of $\underline{M}^{-1}\underline{N}$ are less than 1 in absolute value if (and only if) the invertibility condition holds.

Proof

Observe that M can be written as $\begin{bmatrix}\underline{M} & & & \\ \underline{N} & \underline{M} & & \\ & \underline{N} & \underline{M} & \\ & & \ddots & \\ & & & \ddots\end{bmatrix}$. Direct verification gives the result. For the eigenvalues, see theorem 2.1. □

In view of (2.4), where the correction part of the covariance matrix was introduced, we will now give a lemma regarding the asymptotic behaviour of the trace of this part.

Lemma 5.2

Under the assumption that the zeros of the associated polynomial of the MA part lie all outside the unit circle $\mathbf{tr}(PM^{-1}N-Q)(\underline{P}^T\underline{P}-\underline{QQ}^T)^{-1}(PM^{-1}N-Q)^T$ is bounded for increasing T.

Proof

The problem is of course $PM^{-1}N-Q \in \mathbb{R}^{Txr}$, which becomes larger and larger when T grows. We start by rewriting this expression. As $N=\begin{bmatrix}\underline{N}\\ 0\end{bmatrix}$ we write, in view of lemma 5.1, $M^{-1}N=-\begin{bmatrix}-\underline{M}^{-1}\underline{N}\\ (-\underline{M}^{-1}\underline{N})^2\\ (-\underline{M}^{-1}\underline{N})^3\\ \vdots\\ \vdots\end{bmatrix}$, where $\underline{M}^{-1}\underline{N} \in \mathbb{R}^{rxr}$.

Since $P = \begin{bmatrix} \underline{P} & & & \\ \underline{Q} & \underline{P} & & \\ & \underline{Q} & \underline{P} & \\ & & \ddots & \\ & & & \ddots \end{bmatrix}$,

$$PM^{-1}N\text{-}Q = \begin{bmatrix} -\underline{P}(-\underline{M}^{-1}\underline{N})-\underline{Q} \\ -\underline{Q}(-\underline{M}^{-1}\underline{N})-P(-\underline{M}^{-1}\underline{N})^2 \\ -\underline{Q}(-\underline{M}^{-1}\underline{N})^2-\underline{P}(-\underline{M}^{-1}\underline{N})^3 \\ \vdots \end{bmatrix} = \begin{bmatrix} (\underline{PM}^{-1}\underline{N}\text{-}\underline{Q}) \\ (\underline{PM}^{-1}\underline{N}\text{-}\underline{Q})(-\underline{M}^{-1}\underline{N}) \\ (\underline{PM}^{-1}\underline{N}\text{-}\underline{Q})(-\underline{M}^{-1}\underline{N})^2 \\ \vdots \end{bmatrix} = \begin{bmatrix} A \\ AB \\ AB^2 \\ \vdots \end{bmatrix},$$

where $A = \underline{PM}^{-1}\underline{N}\text{-}\underline{Q} \in \mathbb{R}^{rxr}$ and $B = -\underline{M}^{-1}\underline{N} \in \mathbb{R}^{rxr}$ are introduced to simplify the notation. Write $\underline{D} = \underline{P}^T\underline{P}\text{-}\underline{QQ}^T$, $B^0 = I$, $n = T/r$ and we get

$$\mathrm{tr}(PM^{-1}N\text{-}Q)(\underline{P}^T\underline{P}\text{-}\underline{QQ}^T)^{-1}(PM^{-1}N\text{-}Q)^T =$$

$$= \mathrm{tr}(AB^0\underline{D}^{-1}(B^T)^0A^T + AB\underline{D}^{-1}B^TA^T + AB^2\underline{D}^{-1}(B^T)^2A^T + .. + AB^n\underline{D}^{-1}(B^T)^nA^T)$$

$$= \sum_{k=0}^{n} \mathrm{tr}(B^k\underline{D}^{-1}(B^T)^kA^TA)$$

$$= \sum_{k=0}^{n} vec(A^TA)^T vec(B^k\underline{D}^{-1}(B^T)^k)$$

$$= \sum_{k=0}^{n} (vec(A^TA))^T(B^k \otimes B^k) vec(\underline{D}^{-1})$$

$$= (vec(A^TA))^T \sum_{k=0}^{n} (B^k \otimes B^k) vec(\underline{D}^{-1}).$$

Both A^TA and $\underline{D}^{-1}$ are independent of n. The eigenvalues of $B^k \otimes B^k$ are $(\lambda_i\lambda_j)^k$, i,j=1,..r where λ_i is an eigenvalue of B. Hence $|\lambda_i\lambda_j| < 1$ as $|\lambda_i| < 1$ because of the assumption and theorem 2.1. Therefore the sum is bounded, which concludes the proof. □

When T becomes sufficiently large the influence of the 'starting' values of $\dot{\varepsilon}$ and $\dot{v}$, as defined in section 2.2, will become less and less important. A similar proof is given by Zinde-Walsh (Zinde-Walsh, 1988). She derives the difference between the exact covariance matrix and its infinite form as she calls it; the difference is a matrix of low rank. Here we give an alternative and much shorter proof; we show that the role

the value of the determinant still depends on N and Q. The determinant of the AR case is most illustrative: it depends only on the (p×p) matrices $\underline{P}$ and $\underline{Q}$ and is independent of T. On the other hand, we will show that the value of the determinant is bounded in all cases. Hence its value to the power 1/T will tend to one. That is the reason why the determinant can be disregarded.
To illustrate what happens if T becomes large it may be more clear if we write V as in (2.4): $W+(N-MP^{-1}Q)(\underline{P}^T\underline{P}-\underline{Q}\underline{Q}^T)^{-1}(N-MP^{-1}Q)^T$, where

$$W= P^{-1}M(P^{-1}M)^T. \tag{5.3}$$

Here W can be regarded as the truncated form of V, while the remaining part can be seen as the correction part of the covariance matrix, as stated in section 2.4. We want to show that asymptotically only W is important in estimating β and the ARMA parameters. However, as said before, the determinant of V does not reduce to the determinant of W (which is one). This will be studied in the section 5.4.

5.3 The truncated form of GLS estimators

In this section we will study the behavior of the generalised least square estimates for β and σ^2. The main result is stated in

Theorem 5.1
Let σ^2V be the ARMA covariance matrix and W as defined in (5.3), $\hat{\beta}$ the Aitken estimator based on V and $\tilde{\beta}$ Aitken the estimator based on W. Then
1. $\text{plim}\ (\hat{\sigma}^2-\tilde{\sigma}^2)=0$, where $\hat{\sigma}^2= \varepsilon^TV^{-1}\varepsilon/T$ and $\tilde{\sigma}^2= \varepsilon^TW^{-1}\varepsilon/T$,
2. If $\lim\limits_{T\to\infty} (X^TW^{-1}X/T)^{-1}$ exists, then $\text{plim}\ \sqrt{T}(\hat{\beta}-\tilde{\beta})=0$.

Proof
1. From (2.4) and (5.3) we get $T\cdot(\hat{\sigma}^2-\tilde{\sigma}^2)=\varepsilon^TV^{-1}\varepsilon -\varepsilon^TW^{-1}\varepsilon=$
$\varepsilon^TP^{-1}(PN-MQ)(\underline{P}^T\underline{P}-\underline{Q}\underline{Q}^T)^{-1}(PN-MQ)^TP^{-T}\varepsilon$, which is positive as $\underline{P}^T\underline{P}-\underline{Q}\underline{Q}^T$ is a positive definite matrix.
Straightforward algebra and using $E\varepsilon\varepsilon^T=\sigma^2V$ gives:

$$\begin{aligned}
&E(|\varepsilon^TW^{-1}\varepsilon -\varepsilon^TV^{-1}\varepsilon|)=\\
&\quad= E(\varepsilon^TW^{-1}\varepsilon-\varepsilon^TV^{-1}\varepsilon)\\
&\quad= \mathrm{tr}E(W^{-1}\varepsilon\varepsilon^T-V^{-1}\varepsilon\varepsilon^T)
\end{aligned}$$

$$= \sigma^2 \mathbf{tr}(W^{-1}V - I_T)$$
$$= \sigma^2 \mathbf{tr}(W^{-1}(W+(N-MP^{-1}Q)(\underline{P}^T\underline{P}-\underline{Q}\underline{Q}^T)^{-1}(N-MP^{-1}Q)^T)-I_T)$$
$$= \sigma^2 \mathbf{tr}(P^{-1}M(P^{-1}M)^T)^{-1}(N-MP^{-1}Q)(\underline{P}^T\underline{P}-\underline{Q}\underline{Q}^T)^{-1}(N-MP^{-1}Q)^T$$
$$= \sigma^2 \mathbf{tr}(PM^{-1}N-Q)(\underline{P}^T\underline{P}-\underline{Q}\underline{Q}^T)^{-1}(PM^{-1}N-Q)^T.$$

In view of lemma 5.2 this expression divided by T vanishes for $T \to \infty$. □

2. The second part needs more manipulation. Write the difference between $\hat{\beta}$ and $\tilde{\beta}$ as a function of the error ε.

$\hat{\beta} = (X^TV^{-1}X)^{-1}XV^{-1}y = (X^TV^{-1}X)^{-1}XV^{-1}(X\beta+\varepsilon) = \beta + (X^TV^{-1}X)^{-1}XV^{-1}\varepsilon$

and in the same way

$\tilde{\beta} = \beta + (X^TW^{-1}X)^{-1}XW^{-1}\varepsilon.$

Hence we get for their difference $\hat{\beta} - \tilde{\beta} = Z\varepsilon$, with

$$Z = (X^TV^{-1}X)^{-1}XV^{-1} - (X^TW^{-1}X)^{-1}XW^{-1}, \tag{5.4}$$

$Z \in \mathbb{R}^{k\mathrm{x}T}$. Observe that $E\|Z\varepsilon\|^2 = \mathbf{tr}\ EZ\ \varepsilon\varepsilon^TZ^T = \sigma^2\mathbf{tr}ZVZ^T$. Now plim $\sqrt{T}\ (\hat{\beta}-\tilde{\beta}) =$ plim $\sqrt{T}Z\varepsilon = 0$, if $\lim_{T\to\infty} \sqrt{T}\ E\|Z\varepsilon\|^2 = 0$. We will show, that $\mathbf{tr}Z^TVZ$ has an upper limit similar to the expression in lemma 5.2.

First, however, we define some symbols we will use in this section. Introduce

$$A = N - P^{-1}MQ \in \mathbb{R}^{T\mathrm{x}r}, \tag{5.5}$$
$$\underline{D} = \underline{P}^T\underline{P} - \underline{Q}\underline{Q}^T \in \mathbb{R}^{r\mathrm{x}r}, \tag{5.6}$$
$$H = A^TW^{-1}A + \underline{D} \in \mathbb{R}^{r\mathrm{x}r} \tag{5.7}$$

and

$$G = A^TW^{-1}X(X^TW^{-1}X)^{-1}X^TW^{-1}A - H \in \mathbb{R}^{r\mathrm{x}r}. \tag{5.8}$$

Then $V = W + A\underline{D}^{-1}A^T$, using (5.5) and (5.6). This gives for the inverse of V, using a well known formula for the inverse of the sum of two matrices and (5.7),

$$V^{-1} = (W + A\underline{D}^{-1}A^T)^{-1}$$
$$= W^{-1} - W^{-1}A(A^TW^{-1}A + \underline{D})^{-1}A^TW^{-1}$$

or

$$V^{-1} = W^{-1} - W^{-1}AH^{-1}A^TW^{-1}. \tag{5.9}$$

Next express $(X^TV^{-1}X)^{-1}$ in W and A:

$$(X^TV^{-1}X)^{-1} = (X^TW^{-1}X - X^TW^{-1}AH^{-1}A^TW^{-1}X)^{-1}$$
$$= (X^TW^{-1}X)^{-1} - (X^TW^{-1}X)^{-1}X^TW^{-1}A(A^TW^{-1}X$$
$$(X^TW^{-1}X)^{-1}X^TW^{-1}A - H)^{-1}A^TW^{-1}X(X^TW^{-1}X)^{-1}$$

or using (5.8),

$$(X^TV^{-1}X)^{-1} = (X^TW^{-1}X)^{-1}-(X^TW^{-1}X)^{-1}X^TW^{-1}AG^{-1}A^TW^{-1}X(X^TW^{-1}X)^{-1}. \qquad (5.10)$$

Note that G is negative:

$$\begin{aligned} G &= A^TW^{-1}X(X^TW^{-1}X)^{-1}X^TW^{-1}A-H \\ &= A^TW^{-1}X(X^TW^{-1}X)^{-1}X^TW^{-1}A-A^TW^{-1}A-\underline{D} \\ &= A^T(W^{-1}X(X^TW^{-1}X)^{-1}X^TW^{-1}A-W^{-1})A-\underline{D} \\ &= -A^T(W^{-1}-W^{-1}X(X^TW^{-1}X)^{-1}X^TW^{-1})A-\underline{D}. \end{aligned}$$

Furthermore we need the expression $W^{-1}VW^{-1}$ rewritten in A and W:

$$W^{-1}VW^{-1} = W^{-1}(W+A\underline{D}^{-1}A^T)W^{-1}$$

or

$$W^{-1}VW^{-1} = W^{-1}+W^{-1}A\underline{D}^{-1}A^TW^{-1}. \qquad (5.11)$$

Finally use (5.4) to express ZVZ^T in X and W:

$$\begin{aligned} ZVZ^T &= ((X^TV^{-1}X)^{-1}XV^{-1}-(X^TW^{-1}X)^{-1}XW^{-1})V((X^TV^{-1}X)^{-1}XV^{-1}-(X^TW^{-1}X)^{-1}XW^{-1})^T \\ &= (X^TV^{-1}X)^{-1}XV^{-1}VV^{-1}X^T(X^TV^{-1}X)^{-1}-(X^TV^{-1}X)^{-1}XV^{-1}VW^{-1}X^T(X^TW^{-1}X)^{-1}- \\ &\quad (X^TW^{-1}X)^{-1}XW^{-1}VV^{-1}X^T(X^TV^{-1}X)^{-1}+(X^TW^{-1}X)^{-1}XW^{-1}VW^{-1}X^T(X^TW^{-1}X)^{-1} \\ &= -(X^TV^{-1}X)^{-1}+(X^TW^{-1}X)^{-1}XW^{-1}VW^{-1}X^T(X^TW^{-1}X)^{-1}. \end{aligned}$$

Use (5.10) and (5.11) to replace $(X^TV^{-1}X)^{-1}$ and $W^{-1}VW^{-1}$:

$$\begin{aligned} ZVZ^T &= -(X^TW^{-1}X)^{-1}+(X^TW^{-1}X)^{-1}X^TW^{-1}AG^{-1}A^TW^{-1}X(X^TW^{-1}X)^{-1}+ \\ &\quad (X^TW^{-1}X)^{-1}X(W^{-1}+W^{-1}A\underline{D}^{-1}A^TW^{-1})X^T(X^TW^{-1}X)^{-1} \\ &= -(X^TW^{-1}X)^{-1}+(X^TW^{-1}X)^{-1}X^TW^{-1}AG^{-1}A^TW^{-1}X(X^TW^{-1}X)^{-1}+ \\ &\quad (X^TW^{-1}X)^{-1}+(X^TW^{-1}X)^{-1}XW^{-1}A\underline{D}^{-1}A^TW^{-1}X^T(X^TW^{-1}X)^{-1} \\ &= (X^TW^{-1}X)^{-1}X^TW^{-1}A(G^{-1}+\underline{D}^{-1})A^TW^{-1}X(X^TW^{-1}X)^{-1} \\ &< (X^TW^{-1}X)^{-1}X^TW^{-1}A\underline{D}^{-1}A^TW^{-1}X(X^TW^{-1}X)^{-1}, \end{aligned}$$

because G is definite negative.

Since we assume that the limit of the inverse of $X^TW^{-1}X/T = X^T(P^{-1}M(P^{-1}M)^T)^{-1}X/T = (M^{-1}PX/\sqrt{T})^T(M^{-1}PX/\sqrt{T})$ exists, $\lim_{T\to\infty} (M^{-1}PX/\sqrt{T})$ has to exist. Write $U = \left(\frac{X^TW^{-1}X}{T}\right)^{-1}\left(\frac{(M^{-1}PX)^T}{\sqrt{T}}\right)$. From the definition of W (5.3) and of A (5.5) we have $\frac{X^TW^{-1}A}{T} = \frac{(M^{-1}PX)^T\ M^{-1}P(N-P^{-1}MQ)}{T} = \frac{(M^{-1}PX)^T}{\sqrt{T}}\frac{PM^{-1}N-Q}{\sqrt{T}}$. Hence

$$\lim_{T\to\infty} \sqrt{T}\ \mathrm{tr} ZVZ^T < \sqrt{T}\ \mathrm{tr} \lim_{T\to\infty} \left(\frac{X^TW^{-1}X}{T}\right)^{-1} \left(\frac{X^TW^{-1}A}{T}\right) \underline{D}^{-1} \left(\frac{A^TW^{-1}X}{T}\right) \left(\frac{X^TW^{-1}X}{T}\right)^{-1}$$

$$< \sqrt{T}\ \mathrm{tr} \lim_{T\to\infty} U\ \{(PM^{-1}N-Q)/\sqrt{T}\}\underline{D}^{-1}\{(PM^{-1}N-Q)^T/\sqrt{T}\}\ U^T$$

$$< \mathrm{tr} \lim_{T\to\infty} U \lim_{T\to\infty} \frac{(PM^{-1}N-Q)(\underline{P}^T\underline{P}-\underline{QQ}^T)^{-1}(PM^{-1}N-Q)^T}{\sqrt{T}} \lim_{T\to\infty} U^T$$

$= 0$, in view of lemma 5.2. □

5.4 The limit of the determinant

From the definition of W, equation (5.3), we conclude $|W| = 1$ as W consists of lower band matrices with ones on the main diagonal. But, while $(\varepsilon^TV^{-1}\varepsilon - \varepsilon^TW^{-1}\varepsilon)/T$ tends to zero for $T \to \infty$, the determinant of V will not tend to that of W. In theorem 5.2. we state the result for the asymptotic form of $|V|$.

Theorem 5.2

Let

$$V^* = I_r + (\underline{P}^T\underline{P}-\underline{QQ}^T)^{-1}(\underline{PN}-\underline{MQ})^T(\underline{M}^T\underline{M}-\underline{NN}^T)^{-1}(\underline{PN}-\underline{MQ}). \tag{5.12}$$

If the invertibility condition for the MA part holds, $\lim_{T\to\infty}|V| = |V^*|$.

Proof

As M and P are lower band matrices we have $|M| = |P| = 1$. The determinant of the TxT matrix V is given in (2.8): $|I_r + \underline{D}^{-1}(\underline{PN}-\underline{MQ})^T M^{*T}M^*(\underline{PN}-\underline{MQ})|$, where $M^* \in \mathbb{R}^{T\times r}$ consists of the first r columns of M^{-1}. Hence, in view of (5.1)and writing $n = T/r$:

$$M^* = \begin{bmatrix} \underline{M}^{-1} \\ (-\underline{M}^{-1}\underline{N})\ \underline{M}^{-1} \\ (-\underline{M}^{-1}\underline{N})^2\underline{M}^{-1} \\ \cdot \\ \cdot \end{bmatrix} \text{ and } M^{*T}M^* = \sum_{i=1}^{n} (\underline{M}^{-1})^T((\underline{M}^{-1}\underline{N})^T)^{i-1}(\underline{M}^{-1}\underline{N})^{i-1}(\underline{M}^{-1}).$$

Let this sum be S. Then

$$\underline{M}^TS\underline{M} = \sum_{i=1}^{n} ((\underline{M}^{-1}\underline{N})^T)^{i-1}(\underline{M}^{-1}\underline{N})^{i-1}$$

$$= I_r + \sum_{i=2}^{n} \underline{N}^T(\underline{M}^{-1})^T((\underline{M}^{-1}\underline{N})^T)^{i-2}(\underline{M}^{-1}\underline{N})^{i-2}\underline{M}^{-1}\underline{N}$$

$$= I_r + \underline{N}^T\{\sum_{i=1}^{n}(\underline{M}^{-1})^T((\underline{M}^{-1}\underline{N})^T)^{i-1}(\underline{M}^{-1}\underline{N})^{i-1}\underline{M}^{-1}\}\underline{N}$$

$$= I_r + \underline{N}^T S\underline{N}.$$

This matrix equation, with S as unknown, has a solution if $\underline{M}\otimes\underline{M}-\underline{N}\otimes\underline{N}$ is not singular. Its determinant is $|\underline{M}\otimes\underline{M}-\underline{N}\otimes\underline{N}| = |I_{r^2}-\underline{N}\otimes\underline{N}(\underline{M}\otimes\underline{M})^{-1}|\,|\underline{M}\otimes\underline{M}| = |I_{r^2}-\underline{NM}^{-1}\otimes\underline{NM}^{-1}|$. As the eigenvalues of $\underline{NM}^{-1}$ are smaller than 1 in absolute value, the determinant can not be zero. Its solution S is equal to $(\underline{M}^T\underline{M}-\underline{NN}^T)^{-1}$ or $(\underline{MM}^T-\underline{N}^T\underline{N})^{-1}$, which is positive definite if the invertibility condition 1.4 is met (see section 2.4). These expressions are equivalent as $\underline{M}^T\underline{M}+\underline{N}^T\underline{N} = \underline{MM}^T+\underline{NN}^T$: an equality which is shown easily by comparing the products of the commuting matrices $\begin{bmatrix}\underline{N} & \underline{M}\\ 0 & \underline{N}\end{bmatrix}^T$ and $\begin{bmatrix}\underline{M} & 0\\ \underline{N} & \underline{M}\end{bmatrix}$. Then

$S = (\underline{MM}^T-\underline{N}^T\underline{N})^{-1} \Rightarrow \underline{M}^T S\underline{M} = (I_r-\underline{M}^{-1}\underline{N}^T\underline{N}(\underline{M}^{-1})^T)^{-1}$,

$S = (\underline{M}^T\underline{M}-\underline{NN}^T)^{-1} \Rightarrow \underline{N}^T S\underline{N} = (\underline{N}^{-1}\underline{M}^T\underline{M}(\underline{N}^{-1})^T-I_r)^{-1} = ((\underline{M}^{-1}\underline{N}^T\underline{N}(\underline{M}^{-1})^T)^{-1}-I_r)^{-1}$, where the (commuting) matrices $\underline{M}^{-1}$ and $\underline{N}^T$ are interchanged. Direct verification concludes the proof. □

For the pure AR-case we have $\underline{M}=I$ and $\underline{N}=0$. The determinant becomes $|I_p+(\underline{P}^T\underline{P}-\underline{QQ}^T)^{-1}\underline{QQ}^T| = |(\underline{P}^T\underline{P}-\underline{QQ}^T)^{-1}[\underline{P}^T\underline{P}-\underline{QQ}^T+\underline{Q}^T\underline{Q}]| = |(\underline{P}^T\underline{P}-\underline{QQ}^T)^{-1}\underline{PP}^T| = |(\underline{P}^T\underline{P}-\underline{QQ}^T)^{-1}|$. For the pure MA-case we get a similar result: $|I_q+\underline{N}^T(\underline{M}^T\underline{M}-\underline{NN}^T)^{-1}\underline{N}| = |\underline{M}^T(\underline{M}^T\underline{M}-\underline{NN}^T)^{-1}\underline{M}| = |(\underline{M}^T\underline{M}-\underline{NN}^T)^{-1}|$.

From theorem 5.2 we conclude that the determinant of the ARMA covariance matrix tends to a value that is independent of T. It only depends on the ARMA parameters. Hence we conclude $\lim_{T\to\infty}|V|^{1/T} = \lim_{T\to\infty}|V^*|^{1/T} = 1$.

5.5 Conclusion

In this chapter we showed that is justified to disregard the starting values in the ARMA error structure. If the number of observations is large enough a more simple method can be used to estimate the ARMA parameters. This leads to the so called *Conditional Least Squares* method (CLS) which is widely used, as no closed form expression for the ARMA covariance matrix was available. The theorems given in this chapter can be regarded as a theoretical base for the CLS approach, which will be the subject of the next chapter.

VI CONDITIONAL LEAST SQUARES

6.1 CLS and MD estimation

When the number of observation of observations is large enough the part of the likelihood function which consists of the determinant to the power T^{-1} tends to 1, as we concluded in the previous chapter. At the same time the correction matrix, as proven in section 5.2, has a diminishing influence on the value of the likelihood function. Hence a simplification of the likelihood function is obvious. We can choose from several ways to do so. For the quadratic part we can take either the exact form of the covariance matrix or its truncated form, the one without the correction matrix. For the determinant part we have three alternatives: its exact form, its limiting form or to neglect it completely. Of these six possibilities the exact form is treated in chapter 4. Of the remaining five alternatives two have an exact expression for one and a non-exact expression for the other component. It would not be a consistent way of thought to choose for one of these. The last three alternatives are the so called truncated form (with truncated expressions for both determinant and quadratic form), the Minimum Distance (MD) form (with the exact covariance matrix, but without determinant part) and the Conditional Least Squares (CLS) form (with the covariance matrix in truncated form, without determinant).
There are good reasons *not* to choose for the truncated form. The first reason to neglect it is the small influence the determinant part has on the likelihood function. If the dimensions of the covariance matrix are large enough, the determinant of the covariance matrix approaches to the determinant of the truncated form, which depends only on ARMA parameters and not on the number of rows and columns. As the number of parameters is fixed, its value to the power T^{-1} will (rapidly) tend to 1. A second reason is the complexity of the asymptotic form. In the next section we will give expressions for the first and second differential of this matrix. The computational burden is not justified by the gain in precision.

Thus we can choose between the MD and the CLS method. The most obvious reason to employ CLS is its widely spread use, see *e.g.* Harvey (1981, p. 12-15), who regards CLS estimators as *adequate approximations to the exact ML estimators* (p. 126), however without a formal justification. Hence we neglect the determinant and use $\sigma^2 MP^{-1}(MP^{-1})^T$ as covariance matrix. In the literature this approach is also known as 'pseudo-asymptotic' (Judge, 1984, p.308). In section 2.2 we have rewritten (1.2) as $\begin{bmatrix} Q & P \end{bmatrix}\begin{bmatrix} \bar{\varepsilon} \\ \varepsilon \end{bmatrix} = \begin{bmatrix} N & M \end{bmatrix}\begin{bmatrix} \bar{v} \\ v \end{bmatrix}$. Putting $\bar{\varepsilon}$ and $\bar{v}$ equal to zero gives $P\varepsilon = Mv$, or $\varepsilon = P^{-1}Mv$. The corresponding covariance matrix $E\varepsilon\varepsilon^T$ is $\sigma^2 P^{-1}M(P^{-1}M)^T$ under the condition that $\bar{\varepsilon} = \bar{v} = 0$: hence the name conditional least squares.
Therefore we conclude from (1.12) that we have to minimise

$$S^* = e^T W^{-1} e, \tag{6.1}$$

where W is as defined in (5.3). We will call S^* the Conditional Least Squares function, or shortly the CLS function.

6.2 Determinant of the asymptotic form and its differentials

Now we will have a closer look at the asymptotic form of the determinant, as given in (5.12). We simplify the notation by writing $\underline{D} = \underline{P}^T\underline{P}\text{-}\underline{Q}\underline{Q}^T$, $\Phi = \underline{P}\underline{N}\text{-}\underline{M}\underline{Q}$ and $\underline{E} = \underline{M}^T\underline{M}\text{-}\underline{N}\underline{N}^T$, which gives $|V^*| = |I_r + \underline{D}^{-1}\Phi^T\underline{E}^{-1}\Phi| = |\underline{D}^{-1}||\underline{D} + \Phi^T\underline{E}^{-1}\Phi| = |\underline{D}^{-1}||\underline{E}^{-1}||\Psi|$, where $\Psi = \begin{bmatrix} \underline{E} & -\Phi \\ \Phi^T & \underline{D} \end{bmatrix}$. The matrices $\underline{D}$ and $\underline{E}$ are symmetric, Φ and Ψ are not.

For the pure AR case M = I, N = 0, $\Phi = -Q$ and $|\Psi| = \det\begin{bmatrix} I & Q \\ -Q^T & \underline{D} \end{bmatrix} = |\underline{D} + Q^TQ| = |PP^T| = 1$, with $\mathbf{d}\Psi = 0$ and $V^* = \underline{D}^{-1}$. For the pure MA case $\underline{D} = I_r$, $\Phi = N$, $\Psi = \det\begin{bmatrix} \underline{E} & -N \\ N^T & I \end{bmatrix} = |\underline{E} + N^TN| = |\underline{M}^T\underline{M}| = 1$ and $|V^*|$ reduces to $\underline{E}^{-1}$. Because $\underline{D}$ and $\underline{E}$ are solutions to the covariance equation, the elements (i,i+k) of both $\underline{D}^{-1}$ and $\underline{E}^{-1}$ are independent of i, or, the diagonals of the matrices have the same elements.
Then $\mathbf{d}\log|V^*| = \mathbf{d}\{-\log|\underline{D}| - \log|\underline{E}| + \log|\Psi|\} = -\mathbf{tr}\underline{D}^{-1}\mathbf{d}\underline{D} - \mathbf{tr}\underline{E}^{-1}\mathbf{d}\underline{E} + \mathbf{tr}\Psi^{-1}\mathbf{d}\Psi$.

For the inverse of Ψ we will use

$$\Psi^{-1} = \begin{bmatrix} \Psi_a & \Psi_b^T \\ -\Psi_b & \Psi_c \end{bmatrix} = \begin{bmatrix} (\underline{E}+\Phi\underline{D}^{-1}\Phi^T)^{-1} & (\underline{D}\Phi^{-1}\underline{E}+\Phi^T)^{-1} \\ -(\underline{E}\Phi^{-T}\underline{D}+\Phi)^{-1} & (\underline{D}+\Phi^T\underline{E}^{-1}\Phi)^{-1} \end{bmatrix},$$

where Ψ_a and Ψ_c are symmetric.
Hence,

$$\mathbf{d}\log|V^*| = -\mathbf{tr}\underline{E}^{-1}\mathbf{d}\underline{E}-\mathbf{tr}\underline{D}^{-1}\mathbf{d}\underline{D}+\mathbf{tr}\begin{bmatrix} \Psi_a & \Psi_b^T \\ -\Psi_b & \Psi_c \end{bmatrix}\begin{bmatrix} \mathbf{d}\underline{E} & -\mathbf{d}\Phi \\ \mathbf{d}\Phi^T & \mathbf{d}\underline{D} \end{bmatrix}$$

$$= -\mathbf{tr}\underline{E}^{-1}\mathbf{d}\underline{E}-\mathbf{tr}\underline{D}^{-1}\mathbf{d}\underline{D}+\mathbf{tr}\{\Psi_a\mathbf{d}\underline{E}+\Psi_b^T\mathbf{d}\Phi^T+\Psi_b\mathbf{d}\Phi+\Psi_c\mathbf{d}\underline{D}\}$$

$$= -\mathbf{tr}\underline{E}^{-1}\mathbf{d}\underline{E}-\mathbf{tr}\underline{D}^{-1}\mathbf{d}\underline{D}+\mathbf{tr}\Psi_a\mathbf{d}\underline{E}\ +2\mathbf{tr}\Psi_b\mathbf{d}\Phi\ +\mathbf{tr}\Psi_c\mathbf{d}\underline{D}.$$

The corresponding trace expression for α and ϑ are
$\mathbf{d}_\alpha\log|V^*| = -\mathbf{tr}\underline{E}^{-1}\mathbf{d}\underline{E}\ +\mathbf{tr}(\underline{E}+\Phi\underline{D}^{-1}\Phi^T)^{-1}\mathbf{d}\underline{E}\ +2\mathbf{tr}(\underline{E}\Phi^{-T}\underline{D}+\Phi)^{-1}\mathbf{d}_\alpha\Phi$ and
$\mathbf{d}_\vartheta\log|V^*| = -\mathbf{tr}\underline{D}^{-1}\mathbf{d}\underline{D}\ +\mathbf{tr}(\underline{D}+\Phi^T\underline{E}^{-1}\Phi)^{-1}\mathbf{d}\underline{D}\ +2\mathbf{tr}(\underline{E}\Phi^{-T}\underline{D}+\Phi)^{-1}\mathbf{d}_\vartheta\Phi$
which shows how complicated the differentials of the asymptotic form are. Even worse as can be expected is the second differential:
$\mathbf{d}^2\log|V^*| = \mathbf{tr}\underline{D}^{-1}\mathbf{d}\underline{D}\underline{D}^{-1}\mathbf{d}\underline{D}-\mathbf{tr}\underline{D}^{-1}\mathbf{d}^2\underline{D}+\mathbf{tr}\underline{E}^{-1}\mathbf{d}\underline{E}\underline{E}^{-1}\mathbf{d}\underline{E}-\mathbf{tr}\underline{E}^{-1}\mathbf{d}^2\underline{E}-\mathbf{tr}\Psi^{-1}\mathbf{d}\Psi\Psi^{-1}\mathbf{d}\Psi+\mathbf{tr}\Psi^{-1}\mathbf{d}^2\Psi$, where of course several terms exclude each other. More specifically it becomes for the MA-parameters

$$\mathbf{d}_\alpha^2\log|V^*| = -\mathbf{tr}\Psi^{-1}\mathbf{d}_\alpha\Psi\Psi^{-1}\mathbf{d}_\alpha\Psi+\mathbf{tr}\Psi^{-1}\mathbf{d}^2\Psi_\alpha\ +\mathbf{tr}\underline{E}^{-1}\mathbf{d}\underline{E}\underline{E}^{-1}\mathbf{d}\underline{E}-\mathbf{tr}\underline{E}^{-1}\mathbf{d}^2\underline{E} \qquad (6.2)$$

and for the AR-parameters

$$\mathbf{d}_\vartheta^2\log|V^*| = -\mathbf{tr}\Psi^{-1}\mathbf{d}_\vartheta\Psi\Psi^{-1}\mathbf{d}_\vartheta\Psi+\mathbf{tr}\Psi^{-1}\mathbf{d}_\vartheta^2\Psi+\mathbf{tr}\underline{D}^{-1}\mathbf{d}\underline{D}\underline{D}^{-1}\mathbf{d}\underline{D}-\mathbf{tr}\underline{D}^{-1}\mathbf{d}^2\underline{D}. \qquad (6.3)$$

The differentials $\mathbf{d}\underline{E}$ and $\mathbf{d}\underline{D}$ are used without subscript as they only depend on α or ϑ. In the mixed case we get

$$\mathbf{d}_\alpha\mathbf{d}_\vartheta\log|V^*| = -\mathbf{tr}\Psi^{-1}\mathbf{d}_\alpha\Psi\Psi^{-1}\mathbf{d}_\vartheta\Psi+\mathbf{tr}\Psi^{-1}\mathbf{d}_\alpha\mathbf{d}_\vartheta\Psi.$$

Of course the term $\mathbf{tr}\Psi^{-1}\mathbf{d}\Psi\Psi^{-1}\mathbf{d}\Psi$ is most complicated, as there is no simple expression for the differential of $\mathbf{d}\Psi = \begin{bmatrix} \mathbf{d}\underline{E} & -\mathbf{d}\Phi \\ \mathbf{d}\Phi^T & \mathbf{d}\underline{D} \end{bmatrix}$.

The little gain in (theoretical) precision in employing these truncated forms is not in accordance with the very complicated formulas that result from this approach. This is the main reason not to adopt it although its theoretical foundations. We will now concentrate upon the CLS method.

6.3 First derivatives of the CLS function

The differential of (6.1), the function to be minimised, is $\mathbf{d}S = e^T\mathbf{d}(W^{-1})e$, with $W^{-1} = (P^{-1}M(P^{-1}M)^T)^{-1} = P^T(MM^T)^{-1}P = M^{-T}P^TPM^{-1}$, as P and M commute. The differential becomes for the MA part $e^T\mathbf{d}W^{-1}e = -f^T\mathbf{d}(MM^T)f$ with $f = (MM^T)^{-1}Pe$ and for the AR part $e^T\mathbf{d}W^{-1}e = g^T\mathbf{d}(P^TP)g$ with $g = M^{-1}e$, as $W^{-1} = (M^T)^{-1}P^TPM^{-1}$.

Theorem 6.1

Let $\frac{\partial S^*}{\partial\vartheta} = \left(\frac{\partial S^*}{\partial\vartheta_1}, \frac{\partial S^*}{\partial\vartheta_2}, \ldots, \frac{\partial S^*}{\partial\vartheta_p}\right)^T$ be the vector of AR derivatives and

$\frac{\partial S^*}{\partial\alpha} = \left(\frac{\partial S^*}{\partial\alpha_1}, \frac{\partial S^*}{\partial\alpha_2}, \ldots, \frac{\partial S^*}{\partial\alpha_q}\right)^T$ the vector of MA derivatives of the CLS function.

AR part

Let $g = M^{-1}e$, then $\frac{\partial S^*}{\partial\vartheta} = H\vartheta + h$, where element (i,j) of H is

$$h_{i,j} = 2 \sum_{h=1+\max(i,j)}^{T} g_{h-i}g_{h-j}, \tag{6.4}$$

and $h_i = h_{i,0}$.

MA part

Let $f = (MM^T)^{-1}Pe$, then $\frac{\partial S^*}{\partial\alpha} = A\alpha + a$, where element (i,j) of A is

$$a_{i,j} = \sum_{h=1}^{T-\max(i,j)} f_{h+i}f_{h+j} \tag{6.5}$$

and $a_i = a_{i,0}$.

Proof

AR part

The differential of (6.1), the function to be minimised, is $\mathbf{d}S^* = e^T\mathbf{d}(W^{-1})e$, where $W^{-1} = (P^{-1}M(P^{-1}M)^T)^{-1} = P^T(MM^T)^{-1}P = M^{-T}P^TPM^{-1}$, as P and M commute. For the AR part $e^T\mathbf{d}W^{-1}e = g^T\mathbf{d}(P^TP)g$.

From proposition 3.9 we get the derivative

$$g^T\frac{\partial P^TP}{\partial\vartheta_i}g = g^T\sum_{j=0}^{i}(L_{j-i}(0,j)+L^T_{j-i}(0,j)) + \sum_{j=i+1}^{p}(L_{i-j}(0,i)+L^T_{i-j}(0,i))\vartheta_j\}g$$

$$= 2\{\sum_{j=0}^{i} g^TL_{j-i}(0,j)g + \sum_{j=i+1}^{p} g^TL_{i-j}(0,i)g)\}\vartheta_j$$

$$= 2\{\sum_{j=0}^{i}\sum_{h=1+i}^{T} g_{h-i}g_{h-j} + \sum_{j=i+1}^{p}\sum_{h=1+j}^{T} g_{h-j}g_{h-i}\}\vartheta_j,$$

$$= 2\sum_{j=0}^{p}\sum_{h=1+\max(i,j)}^{T} g_{h-i}g_{h-j}\vartheta_j,$$

as $g^TL_{j-i}(0,j)g = \sum_{h=1}^{T-i} g^T\iota_h\iota^T_{h-j+i}g = \sum_{h=1}^{T-i} g_hg_{h-j+i} = \sum_{h=1+i}^{T} g_{h-i}g_{h-j}$ and similarly

$g^TL_{i-j}(0,i) = \sum_{h=1+j}^{T} g_{h-j}g_{h-i}$.

MA part

As $\mathbf{d}S^* = e^T\mathbf{d}(W^{-1})e = e^T\mathbf{d}(P^T(MM^T)^{-1}P)e = -e^TP^T(MM^T)^{-1}\,\mathbf{d}(MM^T)\,(MM^T)^{-1}Pe = f^T\mathbf{d}(MM^T)f$, we get, using proposition 3.11:

$$f^T\frac{\partial MM^T}{\partial\alpha_i}f = f^T\{\sum_{j=0}^{i} L^T_{j-i}(j,0)+L_{j-i}(j,0)\alpha_j + \sum_{j=i+1}^{q} L_{i-j}(i,0)+L^T_{i-j}(i,0)\alpha_j\}f$$

$$= 2\sum_{j=0}^{i} f^TL_{j-i}(j,0)f\alpha_j + 2\sum_{j=i+1}^{q} f^TL_{i-j}(i,0)f\alpha_j$$

$$= 2\sum_{h=1}^{T-\max(i,j)} f_{h+j}f_{h+i}\alpha_j,$$

as $f^TL_{j-i}(j,0)f = \sum_{h=1+j}^{T+j-i} f^T\iota\iota^T_{hh-j+i}f = \sum_{h=1+j}^{T+j-i} f_hf_{h-j+i} = \sum_{h=1}^{T-i} f_{h+j}f_{h+i}$ and in the same way $f^TL_{i-j}(i,0)f = \sum_{h=1}^{T-j} f_{h+j}f_{h+i}$. □

Corollary 6.1 *Pure AR-case*

$\frac{\partial S^*}{\partial \theta} = H\theta + h$, where element (i,j) of H is

$$h_{i,j} = 2 \sum_{h=1+\max(i,j)}^{T} e_{h-i}e_{h-j}, \tag{6.6}$$

and $h_i = h_{i,0}$.

Proof

Obvious as $M = I_T$ and thus $g = e$. □

Remark. These are of course the well known Yule-Walker conditions.

Corollary 6.2 *Pure MA-case*

Let $f^* = (MM^T)^{-1}e$, then $\frac{\partial S^*}{\partial \alpha} = A\alpha + a$, where element (i,j) of A is

$$a_{i,j} = \sum_{h=1}^{T-\max(i-j)} f^*_{h+i}f^*_{h+j} \tag{6.7}$$

and $a_i = a_{i,0}$.

Proof

Obvious from (6.5). □

6.4 Second differential

From the first differential $\mathbf{d}S^* = e^T\mathbf{d}(W^{-1})e$, we get the general form of the second differential:

$$\mathbf{d}^2S^* = e^T\mathbf{d}^2W^{-1}e - 2e^T\mathbf{d}W^{-1}X(X^TW^{-1}X)^{-1}X^T\mathbf{d}W^{-1}e, \tag{6.8}$$

or equivalently, with differentials for W instead of W^{-1}:

$$\mathbf{d}^2S^* = -e^TW^{-1}\mathbf{d}^2WW^{-1}e + 2\,e^TW^{-1}\mathbf{d}W(W^{-1} - W^{-1}X(X^TW^{-1}X)^{-1}X^TW^{-1})\mathbf{d}WW^{-1}e \tag{6.9}$$

For the three specific cases we give the differentials in the next theorem:

Theorem 6.2

Second differential of AR parameters

Let $g = M^{-1}e$, then

$$\mathbf{d}_\vartheta^2 S^* = +2\ g^T\mathbf{d}P^T\mathbf{d}Pg\ -2\ g^T\mathbf{d}(P^TP)M^{-1}X(X^TW^{-1}X)^{-1}X^TM^{-T}\mathbf{d}(P^TP)g. \qquad (6.10)$$

Second differential of MA parameters

Let $f = M^{-T}Pg$, then

$$\mathbf{d}_\alpha^2 S^* =$$
$$-\ 2f\mathbf{d}M\ \mathbf{d}M^Tf\ +2\ f^T\mathbf{d}(MM^T)P^{-T}(W^{-1}\ -W^{-1}X(X^TW^{-1}X)^{-1}X^TW^{-1})P^{-1}\mathbf{d}(MM^T)f. \qquad (6.11)$$

Second differential of mixed part

Let $h = M^{-1}PM^{-1}e = M^{-1}Pg$, then

$$\mathbf{d}^2 S^* =$$
$$-2g^T\mathbf{d}P^T\mathbf{d}Mh - 2f^T\mathbf{d}M\mathbf{d}Pg + 2g^T\mathbf{d}(P^TP)M^{-1}X(X^TW^{-1}X)^{-1}X^TP^T(MM^T)^{-1}\mathbf{d}(MM^T)f. \qquad (6.12)$$

Proof

AR part

From (5.3) and $g = M^{-1}e$ we get $e^T\mathbf{d}_\vartheta W^{-1}e = e^TM^{-T}\mathbf{d}(P^TP)M^{-1}e = g^T\mathbf{d}_\vartheta(P^TP)g = g^T\mathbf{d}_\vartheta P^TPg + g^TP^T\mathbf{d}_\vartheta Pg = 2g^TP^T\mathbf{d}Pg$. As the elements of P are a linear function of ϑ, we get $\mathbf{d}_\vartheta^2 W^{-1} = 2g^T\mathbf{d}P^T\mathbf{d}Pg$. For $\mathbf{d}W^{-1}e$, part of the last term, we use $M^{-T}\mathbf{d}_\vartheta(P^TP)g$. Substitution into (6.9) gives (6.12). □

MA part

In this case we will use (6.9) as we have now expressions for the differential of W instead of W^{-1}. As $\mathbf{d}_\alpha^2 M = 0$, we conclude $\mathbf{d}_\alpha^2(MM^T) = \mathbf{d}_\alpha(\mathbf{d}_\alpha MM^T + M\mathbf{d}_\alpha M^T) = 2\mathbf{d}_\alpha M\mathbf{d}_\alpha M^T$ and thus $\mathbf{d}W = P^{-1}\mathbf{d}_\alpha(MM^T)P^{-T}$ and $\mathbf{d}_\alpha^2 W = 2P^{-1}\mathbf{d}_\alpha M\mathbf{d}_\alpha M^TP^{-T}$. Hence $e^TW^{-1}\mathbf{d}_\alpha WW^{-1}e =$ $e^TP^TM^{-T}M^{-1}P\cdot P^{-1}\mathbf{d}_\alpha(MM^T)P^{-T}\cdot P^TM^{-T}M^{-1}Pe = f^T\mathbf{d}(MM^T)f$. For the second differential we get $e^TW^{-1}\mathbf{d}^2WW^{-1}e = 2e^TP^TM^{-T}M^{-1}P\cdot P^{-1}\mathbf{d}_\alpha M\mathbf{d}_\alpha M^TP^{-T}\cdot P^TM^{-T}M^{-1}Pe = 2f^T\mathbf{d}_\alpha M\mathbf{d}_\alpha Mf$. In the last term of the second differential we write $\mathbf{d}WW^{-1}e$ as $P^{-1}\mathbf{d}(MM^T)P^{-T}P^TM^{-T}M^{-1}Pe = P^{-1}\mathbf{d}(MM^T)f$. □

Mixed part

Starting from (6.8) we need $\mathbf{d}_\alpha W^{-1}e$, $\mathbf{d}_\vartheta W^{-1}e$ and the second differential $e^T\mathbf{d}^2W^{-1}e$. For the first two we have

$$\begin{aligned}\mathbf{d}_\alpha W^{-1}e &= \mathbf{d}_\alpha(P^T(MM^T)^{-1}P)e\\ &= -P^T(MM^T)^{-1}\mathbf{d}(MM^T)(MM^T)^{-1}Pe\\ &= -P^T(MM^T)^{-1}\mathbf{d}(MM^T)f\end{aligned}$$

and

$$\begin{aligned}\mathbf{d}_\vartheta W^{-1}e &= \mathbf{d}_\vartheta(M^{-T}P^TPM^{-1})e\\ &= M^{-T}\mathbf{d}(P^TP)M^{-1}e\\ &= M^{-T}\mathbf{d}(P^TP)g\ .\end{aligned}$$

The second differential can be expressed in several ways depending on the choices regarding transposes and the sequences of symbols. One possibility is the following using symbols already used:

$$\begin{aligned}e^T\mathbf{d}^2W^{-1}e &= e^T\mathbf{d}_\vartheta\mathbf{d}_\alpha(P^TM^{-T}M^{-1}P)e\\ &= -e^T\mathbf{d}_\vartheta(P^TM^{-T}\mathbf{d}M^TM^{-T}M^{-1}P + P^TM^{-T}M^{-1}\mathbf{d}MM^{-1}P)e\\ &= -2e^T\mathbf{d}_\vartheta(P^TM^{-T}M^{-1}\mathbf{d}MM^{-1}P)e\\ &= -2e^T(\mathbf{d}P^TM^{-T}M^{-1}\mathbf{d}MM^{-1}P + P^TM^{-T}M^{-1}\mathbf{d}MM^{-1}\mathbf{d}P)e\\ &= -2e^TM^{-T}\mathbf{d}P^T\ \mathbf{d}MM^{-1}PM^{-1}e\ -\ 2e^TP^TM^{-T}M^{-1}\mathbf{d}M\ \mathbf{d}PM^{-1}e\\ &= -2g^T\mathbf{d}P^T\mathbf{d}Mh-2f^T\mathbf{d}M\mathbf{d}Pg\end{aligned}$$

where $h = M^{-1}PM^{-1}e = M^{-1}Pg$. □

6.5 Second derivatives

From these differentials we derive the second derivatives of the CLS function.

Theorem 6.3

Second derivatives of AR parameters

Let $g = M^{-1}e$, $H_1 = M^{-1}X(X^TW^{-1}X)^{-1}XM^T$ and W as defined in (5.3). Then

$$\frac{\partial^2 S^*}{\partial\vartheta_i\partial\vartheta_j} = 2g^T\frac{\partial P^T}{\partial\vartheta_i}\frac{\partial P}{\partial\vartheta_j}g\ -\ 2g^T\frac{\partial(P^TP)}{\partial\vartheta_i}M^{-1}X(X^TW^{-1}X)^{-1}XM^T\frac{\partial(P^TP)}{\partial\vartheta_j}g, \tag{6.13}$$

with

$$g^T\frac{\partial P^T}{\partial\vartheta_i}\frac{\partial P}{\partial\vartheta_j}g = 2\sum_{h=1+\max(i,j)}^{T} g_{h-i}g_{h-j} \tag{6.13.1}$$

and

$$2g^T\frac{\partial(P^TP)}{\partial\vartheta_i}H_1\frac{\partial(P^TP)}{\partial\vartheta_j}g =$$

$$\Big\{\sum_{k=0}^{i}\sum_{l=0}^{j}\sum_{s=1-k}^{T-i-k}\sum_{t=1-l}^{T-j-l}+\sum_{k=0}^{i}\sum_{l=j+1}^{p}\sum_{s=1-k}^{T-i-k}\sum_{t=1-j}^{T-l-j}+$$

$$\sum_{k=i+1}^{p}\sum_{l=0}^{j}\sum_{s=1-i}^{T-k-i}\sum_{t=1-l}^{T-j-l}+\sum_{k=i+1}^{p}\sum_{l=j+1}^{p}\sum_{s=1-i}^{T-k-i}\sum_{t=1-j}^{T-l-j}\Big\}$$

$$(g_{s+i}g_{t+j}H_1[s+k,t+l]+g_{s+i}g_{t+l}H_1[s+k,t+j]+$$
$$g_{s+k}g_{t+j}H_1[s+i,t+l]+g_{s+k}g_{t+l}H_1[s+i,t+j])\vartheta_k\vartheta_l. \tag{6.13.2}$$

Second derivatives of MA parameters

Let $f=(MM^T)^{-1}Pe$ and $H_2=P^{-T}(W^{-1}-W^{-1}X(X^TW^{-1}X)^{-1}X^TW^{-1})P^{-1}$. Then

$$\frac{\partial^2S^*}{\partial\alpha_i\partial\alpha_j}=-2\ f^T\frac{\partial M}{\partial\alpha_i}\frac{\partial M^T}{\partial\alpha_j}f\ +2f^T\frac{\partial(MM^T)}{\partial\alpha_i}H_2\frac{\partial(MM^T)}{\partial\alpha_j}f \tag{6.14}$$

with

$$f^T\frac{\partial M}{\partial\alpha_i}\frac{\partial M^T}{\partial\alpha_j}f=\sum_{h=1}^{T-\max(i,j)}f_{h+i}f_{h+j} \tag{6.14.1}$$

and

$$f^T\frac{\partial(MM^T)}{\partial\alpha_i}H_2\frac{\partial(MM^T)}{\partial\alpha_j}f=\Big\{\sum_{k=0}^{i}\sum_{l=0}^{j}\sum_{s=1}^{T-i}\sum_{t=1}^{T-j}+\sum_{k=0}^{i}\sum_{l=j+1}^{q}\sum_{s=1}^{T-i}\sum_{t=1}^{T-l}+$$

$$\sum_{k=i+1}^{q}\sum_{l=0}^{j}\sum_{s=1}^{T-k}\sum_{t=1}^{T-j}+\sum_{k=i+1}^{q}\sum_{l=j+1}^{q}\sum_{s=1}^{T-k}\sum_{t=1}^{T-l}\Big\}$$

$$(f_{s+k}f_{t+l}H_2[s+i,t+j]+f_{s+k}f_{t+j}H_2[s+i,t+l]+$$
$$f_{s+i}f_{t+l}H_2[s+k,t+j]+f_{s+i}f_{t+j}H_2[s+k,t+l])\alpha_k\alpha_l. \tag{6.14.2}$$

Mixed derivatives

Let $h=M^{-1}PM^{-1}e$ and $H_3=M^{-1}X(X^TW^{-1}X)^{-1}X^TP^T(MM^T)^{-1}$

$$\frac{\partial^2S^*}{\partial\alpha_i\partial\vartheta_j}=\ -2g^T\frac{\partial P^T}{\partial\vartheta_j}\frac{\partial M}{\partial\alpha_i}h\ -2\ f^T\frac{\partial M}{\partial\alpha_i}\frac{\partial P}{\partial\vartheta_j}g\ +\ 2g^T\frac{\partial(P^TP)}{\partial\vartheta_i}H_3\frac{\partial(MM^T)}{\partial\alpha_j}f \tag{6.15}$$

$$g^T\frac{\partial P^T}{\partial\vartheta_j}\frac{\partial M}{\partial\alpha_i}h=\sum_{i=0}^{q}\sum_{j=0}^{p}\sum_{k=1-\min(i,j)}^{T-i-j}g_{k+i}h_{k+j} \tag{6.15.1}$$

$$f^T\frac{\partial M}{\partial \alpha_i}\frac{\partial P}{\partial \vartheta_j}g = \sum_{i=0}^{q}\sum_{j=0}^{p}\sum_{k=j+1}^{T-i} f_{k+i}g_{k-j} \tag{6.15.2}$$

$$g^T\frac{\partial(P^TP)}{\partial \vartheta_i}H_3\frac{\partial(MM^T)}{\partial \alpha_j}f =$$

$$\{\sum_{k=0}^{i}\sum_{l=0}^{j}\sum_{s=1-k}^{T-i-k}\sum_{t=1}^{T-j} + \sum_{k=0}^{i}\sum_{l=j+1}^{q}\sum_{s=1-k}^{T-i-k}\sum_{t=1}^{T-l} +$$

$$\sum_{k=i+1}^{p}\sum_{l=0}^{j}\sum_{s=1-i}^{T-k-i}\sum_{t=1}^{T-j} + \sum_{k=i+1}^{p}\sum_{l=j+1}^{q}\sum_{s=1-i}^{T-k-i}\sum_{t=1}^{T-l}\}$$

$$(g_{s+k}f_{t+l}H_3[s+i,t+j] + g_{s+k}f_{t+j}H_3[s+i,t+l] +$$

$$g_{s+i}f_{t+l}H_3[s+k,t+j] + g_{s+i}f_{t+j}H_3[s+k,t+l])\vartheta_k\alpha_l. \tag{6.15.3}$$

Proof

Second derivatives of AR parameters

From the differential (6.8) the derivatives follow immediately.

For the expression for $g^T\frac{\partial P^T\partial P}{\partial\vartheta_i\partial\vartheta_j}g$ we refer to the proof of (6.4): it is the first derivative once more differentiated.

The second part, (6.13.2), is found by using proposition 3.9:

$$\frac{\partial(P^TP)}{\partial\vartheta_i}g = \{\sum_{k=0}^{i}(L_{k-i}(0,k) + L^T_{k-i}(0,k)) + \sum_{k=i+1}^{p}(L_{i-k}(0,i) + L^T_{i-k}(0,i))\vartheta_k\}g$$

$$= \{\sum_{k=0}^{i}(L_{k-i}(0,k) + L_{i-k}(i-k,i))g + \sum_{k=i+1}^{p}(L_{i-k}(0,i) + L_{k-i}(k-i,k))g\}\vartheta_k$$

$$= \{\sum_{k=0}^{i}(\sum_{h=1}^{T-i}\iota_h^T\iota_{h-k+i} + \sum_{h=1+i-k}^{T-k}\iota_h^T\iota_{h-i+k})g + \sum_{k=i+1}^{p}(\sum_{h=1}^{T-k}\iota_h^T\iota_{h-i+k} + \sum_{h=1+k-i}^{T-i}\iota_h^T\iota_{h-k+i}g\}\vartheta_k$$

$$= \{\sum_{k=0}^{i}(\sum_{h=1}^{T-i}g_h\iota_{h-k+i} + \sum_{h=1+i-k}^{T-k}g_h\iota_{h-i+k} + \sum_{k=i+1}^{p}(\sum_{h=1}^{T-k}g_h\iota_{h-i+k} + \sum_{h=1+k-i}^{T-i}g_h\iota_{h-k+i})\}\vartheta_k$$

$$= \{\sum_{k=0}^{i}\sum_{h=1-k}^{T-i-k}(g_{h+k}\iota_{h+i} + g_{h+i}\iota_{h+k}) + \sum_{k=i+1}^{p}\sum_{h=1-i}^{T-k-i}(g_{h+i}\iota_{h+k} + g_{h+k}\iota_{h+i})\}\vartheta_k.$$

For the quadratic form we get

$$g^T\frac{\partial(P^TP)}{\partial\vartheta_i}H_1\frac{\partial(P^TP)}{\partial\vartheta_j}g =$$

$$=\Big\{\sum_{k=0}^{i}\sum_{s=1-k}^{T-i-k}(g_{s+k}\iota_{s+i}+g_{s+i}\iota_{s+k}) + \sum_{k=i+1}^{p}\sum_{s=1-i}^{T-k-i}(g_{s+i}\iota_{s+k}+g_{s+k}\iota_{s+i})\Big\}\vartheta_k\ H_1$$
$$\sum_{l=0}^{j}\sum_{t=1-l}^{T-j-l}(g_{t+l}\iota_{t+j}+g_{t+j}\iota_{t+l}) + \sum_{l=j+1}^{p}\sum_{t=1-j}^{T-l-j}(g_{t+j}\iota_{t+l}+g_{t+l}\iota_{t+j})\Big\}\vartheta_l$$

$$=\Big\{\sum_{k=0}^{i}\sum_{l=0}^{j}\sum_{s=1-k}^{T-i-k}\sum_{t=1-l}^{T-j-l}(g_{s+k}\iota_{s+i}+g_{s+i}\iota_{s+k})H_1(g_{t+l}\iota_{t+j}+g_{t+j}\iota_{t+l})+$$
$$\sum_{k=0}^{i}\sum_{l=j+1}^{p}\sum_{s=1-k}^{T-i-k}\sum_{t=1-j}^{T-l-j}(g_{s+k}\iota_{s+i}+g_{s+i}\iota_{s+k})H_1(g_{t+j}\iota_{t+l}+g_{t+l}\iota_{t+j})+$$
$$\sum_{k=i+1}^{p}\sum_{l=0}^{j}\sum_{s=1-i}^{T-k-i}\sum_{t=1-l}^{T-j-l}(g_{s+i}\iota_{s+k}+g_{s+k}\iota_{s+i})H_1(g_{t+l}\iota_{t+j}+g_{t+j}\iota_{t+l})+$$
$$\sum_{k=i+1}^{p}\sum_{l=j+1}^{p}\sum_{s=1-i}^{T-k-i}\sum_{t=1-j}^{T-l-j}(g_{s+i}\iota_{s+k}+g_{s+k}\iota_{s+i})H_1(g_{t+j}\iota_{t+l}+g_{t+l}\iota_{t+j})\Big\}\vartheta_k\vartheta_l$$

$$=\Big\{\sum_{k=0}^{i}\sum_{l=0}^{j}\sum_{s=1-k}^{T-i-k}\sum_{t=1-l}^{T-j-l}+\sum_{k=0}^{i}\sum_{l=j+1}^{p}\sum_{s=1-k}^{T-i-k}\sum_{t=1-j}^{T-l-j}+$$
$$\sum_{k=i+1}^{p}\sum_{l=0}^{j}\sum_{s=1-i}^{T-k-i}\sum_{t=1-l}^{T-j-l}+\sum_{k=i+1}^{p}\sum_{l=j+1}^{p}\sum_{s=1-i}^{T-k-i}\sum_{t=1-j}^{T-l-j}\Big\}$$
$$(g_{s+i}g_{t+j}H_1[s+k,t+l]+g_{s+i}g_{t+l}H_1[s+k,t+j]+$$
$$g_{s+k}g_{t+j}H_1[s+i,t+l]+g_{s+k}g_{t+l}H_1[s+i,t+j])\vartheta_k\vartheta_l. \quad \square$$

Remark. Although this derivation follows the same lines as we used to do before, computationally it is more profitable to compute $\frac{\partial(P^TP)}{\partial\vartheta_i}g$ first:

$$\frac{\partial(P^TP)}{\partial\vartheta_i}g=\Big\{\sum_{k=0}^{i}(L_{k-i}(0,k)+L_{k-i}^T(0,k))\vartheta_k+\sum_{k=i+1}^{p}(L_{i-k}(0,i)+L_{i-k}^T(0,i))\vartheta_k\Big\}g$$
$$=\sum_{k=0}^{i}\mathfrak{g}(i,k)\vartheta_k+\sum_{k=i+1}^{p}\mathfrak{g}(k,i)\vartheta_k$$

with

$$\mathfrak{g}(i,k) = L_{k-i}(0,k)g + L_{k-i}^T(0,k)g$$

$$= \sum_{h=1}^{T-i} (\iota_h \iota_{h-k+i}^T g + (\iota_h \iota_{h-k+i}^T)^T g)$$

$$= \sum_{h=1}^{T-i} (\iota_h g_{h-k+i} + \iota_{h-k+i} g_h)$$

$$= (g_{i-k+1} \ldots g_{T-k} \underbrace{0 \ldots 0}_{\leftarrow i \rightarrow})^T + (\underbrace{0 \ldots 0}_{\leftarrow i-k \rightarrow} g_1 \ldots g_{T-i} \underbrace{0 \ldots 0}_{\leftarrow i \rightarrow})^T.$$

Second derivatives of the MA parameters

The proof follows the same lines as in the case of the AR parameters. However, there is one difference, as the transposed part comes first in the AR case. Hence the result is slightly different. Expression (6.14.1) is simply found by differentiating $f^T \frac{\partial MM^T}{\partial\alpha_i \partial\alpha_j} f$ once more, see (6.5).

For (6.14.2) we first compute

$$\frac{\partial(MM^T)}{\partial\alpha_i} f = \{ \sum_{k=0}^{i} L_{k-i}^T(k,0) + L_{k-i}(k,0)\alpha_k + \sum_{k=i+1}^{q} L_{i-k}(i,0) + L_{i-k}^T(i,0)\alpha_k \} f$$

$$= \{ \sum_{k=0}^{i} L_{i-k}(i,k-i) + L_{k-i}(k,0) + \sum_{k=i+1}^{q} L_{i-k}(i,0) + L_{k-i}(k,k-i) \} f\alpha_k$$

$$= \{ \sum_{k=0}^{i} \sum_{h=1+i}^{T} \iota_h \iota_{h-i+k}^T + \sum_{h=1+k}^{T+k-i} \iota_h \iota_{h+i-k}^T + \sum_{k=i+1}^{q} \sum_{h=1+i}^{T+i-k} \iota_h \iota_{h-i+k}^T + \sum_{h=1+k}^{T} \iota_h \iota_{h+i-k}^T \} f\alpha_k$$

$$= \{ \sum_{k=0}^{i} \sum_{h=1}^{T-i} (\iota_{h+i} f_{h+k} + \iota_{h+k} f_{h+i}) + \sum_{k=i+1}^{q} \sum_{h=1}^{T-k} (\iota_{h+i} f_{h+k} + \iota_{h+k} f_{h+i}) \} \alpha_k.$$

Then (6.14.2) becomes

$$f^T \frac{\partial(MM^T)}{\partial\alpha_i} H_2 \frac{\partial(MM^T)}{\partial\alpha_j} f =$$

$$= \{ \sum_{k=0}^{i} \sum_{s=1}^{T-i} f_{s+k} \iota_{s+i}^T + f_{s+i} \iota_{s+k}^T) + \sum_{k=i+1}^{q} \sum_{s=1}^{T-k} (f_{s+k} \iota_{s+i}^T + f_{s+i} \iota_{s+k}^T) \} \; H$$

$$\{ \sum_{l=0}^{j} \sum_{t=1}^{T-j} (\iota_{t+j} f_{t+l} + \iota_{t+l} f_{t+j}) + \sum_{l=j+1}^{q} \sum_{t=1}^{T-l} (\iota_{t+j} f_{t+l} + \iota_{t+l} f_{t+j}) \} \alpha_k \alpha_l$$

$$=\{\sum_{k=0}^{i}\sum_{l=0}^{j}\sum_{s=1}^{T-i}\sum_{t=1}^{T-j}+\sum_{k=0}^{i}\sum_{l=j+1}^{q}\sum_{s=1}^{T-i}\sum_{t=1}^{T-l}+\sum_{k=i+1}^{q}\sum_{l=0}^{j}\sum_{s=1}^{T-k}\sum_{t=1}^{T-j}+$$
$$\sum_{k=i+1}^{q}\sum_{l=j+1}^{q}\sum_{s=1}^{T-k}\sum_{t=1}^{T-l}\}(f_{s+k}f_{t+l}H[s+i,t+j]+f_{s+k}f_{t+j}H[s+i,t+l]+$$
$$f_{s+i}f_{t+l}H[s+k,t+j]+f_{s+i}f_{t+j}H[s+k,t+l])\alpha_k\alpha_l. \ \square$$

Remark. As in the AR case, we also can form $\frac{\partial MM^T}{\partial\alpha_i}f$ separately:

$$\frac{\partial MM^T}{\partial\alpha_i}f=\{\sum_{k=0}^{i}(L_{k-i}(k,0)+L^T_{k-i}(k,0))\alpha_k+\sum_{k=i+1}^{q}(L_{i-k}(i,0)+L^T_{i-k}(i,0))\alpha_k\}f$$
$$=\sum_{k=0}^{i}(L_{k-i}(k,0)f+L^T_{k-i}(k,0)f)\alpha_k+\sum_{k=i+1}^{q}(L_{i-k}(i,0)f+L^T_{i-k}(i,0)f)\alpha_k$$
$$=\sum_{k=0}^{i}\mathcal{F}(i,k)\alpha_k+\sum_{k=i+1}^{q}\mathcal{F}(k,i)\alpha_k$$

with

$$\mathcal{F}(i,k)=L_{k-i}(k,0)f+L^T_{k-i}(k,0)f$$
$$=\sum_{h=1+k}^{T-i+k}\iota_h\iota^T_{h-k+i}f+(\iota_h\iota^T_{h-k+i})^Tf$$
$$=\sum_{h=1+k}^{T-i+k}\iota_h f_{h-k+i}+\iota_{h-k+i}f_h$$
$$=(\underbrace{0\ldots0}_{\leftarrow k\rightarrow}\ f_{i+1}\ldots f_T\ \underbrace{0\ldots0}_{\leftarrow i-k\rightarrow})^T+(\ \underbrace{0\ldots0}_{\leftarrow i\rightarrow}\ f_{1+k}\ldots f_{T-i+k})^T.$$

Eventually we treat the mixed derivatives of the CLS function.

Mixed derivatives

Equation (6.15.1) follows from

$$g^T\frac{\partial P^T}{\partial\vartheta_j}\frac{\partial M}{\partial\alpha_i}h=g^T\sum_{j=0}^{p}L^T_j(j)(\sum_{i=0}^{q}L_i(i)h$$
$$=\sum_{i=0}^{q}\sum_{j=0}^{p}g^TL_{-j}(0)(L_i(i)h$$

$$= \sum_{i=0}^{q} \sum_{j=0}^{p} g^T L_{i-j}(\max(0,i-j),i) h \alpha_i \vartheta_j$$

$$= \sum_{i=0}^{q} \sum_{j=0}^{p} \sum_{k=1+\max(0,i-j)}^{T-j} g^T \iota_k \iota_{k-i+j}^T h$$

$$= \sum_{i=0}^{q} \sum_{j=0}^{p} \sum_{k=1+\max(0,i-j)}^{T-j} g_k h_{k-i+j}$$

$$= \sum_{i=0}^{q} \sum_{j=0}^{p} \sum_{k=1-\min(i,j)}^{T-i-j} g_{k+i} h_{k+j}.$$

For (6.15.2) we get

$$f^T \frac{\partial M}{\partial \alpha_i} \frac{\partial P}{\partial \vartheta_j} g = f^T \sum_{i=0}^{q} L_i(i) \sum_{j=0}^{p} L_j(j) g$$

$$= f^T (\sum_{i=0}^{q} \sum_{j=0}^{p} L_{i+j}(i+j) g$$

$$= \sum_{i=0}^{q} \sum_{j=0}^{p} \sum_{k=i+j+1}^{T} f^T \iota_k \iota_{k-i-j}^T g$$

$$= \sum_{i=0}^{q} \sum_{j=0}^{p} \sum_{k=j+1}^{T-i} f_{k+i} g_{k-j}.$$

The last expression (6.15.3) becomes

$$g^T \frac{\partial (P^T P)}{\partial \vartheta_i} H_3 \frac{\partial (M M^T)}{\partial \alpha_j} f =$$

$$= \{ \sum_{k=0}^{i} \sum_{s=1-k}^{T-i-k} (g_{s+k}\iota_{s+i} + g_{s+i}\iota_{s+k}) + \sum_{k=i+1}^{p} \sum_{s=1-i}^{T-k-i} (g_{s+i}\iota_{s+k} + g_{s+k}\iota_{s+i}) \}^T \vartheta_k \; H$$

$$\{ \sum_{l=0}^{j} \sum_{t=1}^{T-j} (\iota_{t+j} f_{t+l} + \iota_{t+l} f_{t+j}) + \sum_{l=j+1}^{q} \sum_{t=1}^{T-l} (\iota_{t+j} f_{t+l} + \iota_{t+l} f_{t+j}) \} \alpha_l$$

$$=\Big\{\sum_{k=0}^{i}\ \sum_{s=1-k}^{T-i-k}(g_{s+k}\iota_{s+i}^{T}+g_{s+i}\iota_{s+k}^{T})\ +\sum_{k=i+1}^{p}\ \sum_{s=1-i}^{T-k-i}(g_{s+i}\iota_{s+k}^{T}+g_{s+k}\iota_{s+i}^{T})\Big\}\ H$$
$$\Big\{\sum_{l=0}^{j}\ \sum_{t=1}^{T-j}(\iota_{t+j}f_{t+l}+\iota_{t+l}f_{t+j})+\sum_{l=j+1}^{q}\ \sum_{t=1}^{T-l}(\iota_{t+j}f_{t+l}+\iota_{t+l}f_{t+j})\Big\}\vartheta_k\alpha_l$$

$$=\Big\{\sum_{k=0}^{i}\ \sum_{l=0}^{j}\ \sum_{s=1-k}^{T-i-k}\ \sum_{t=1}^{T-j}(g_{s+k}\iota_{s+i}^{T}+g_{s+i}\iota_{s+k}^{T})\ H\ (\iota_{t+j}f_{t+l}+\iota_{t+l}f_{t+j})+$$
$$\sum_{k=0}^{i}\ \sum_{l=j+1}^{q}\ \sum_{s=1-k}^{T-i-k}\ \sum_{t=1}^{T-l}(g_{s+k}\iota_{s+i}^{T}+g_{s+i}\iota_{s+k}^{T})\ H\ (\iota_{t+j}f_{t+l}+\iota_{t+l}f_{t+j})+$$
$$\sum_{k=i+1}^{p}\ \sum_{l=0}^{j}\ \sum_{s=1-i}^{T-k-i}\ \sum_{t=1}^{T-j}(g_{s+i}\iota_{s+k}^{T}+g_{s+k}\iota_{s+i}^{T})\ H\ (\iota_{t+j}f_{t+l}+\iota_{t+l}f_{t+j})+$$
$$\sum_{k=i+1}^{p}\ \sum_{l=j+1}^{q}\ \sum_{s=1-i}^{T-k-i}\ \sum_{t=1}^{T-l}(g_{s+i}\iota_{s+k}^{T}+g_{s+k}\iota_{s+i}^{T})\ H\ (\iota_{t+j}f_{t+l}+\iota_{t+l}f_{t+j})\Big\}\vartheta_k\alpha_l$$

$$=\Big\{\sum_{k=0}^{i}\ \sum_{l=0}^{j}\ \sum_{s=1-k}^{T-i-k}\ \sum_{t=1}^{T-j}+\sum_{k=0}^{i}\ \sum_{l=j+1}^{q}\ \sum_{s=1-k}^{T-i-k}\ \sum_{t=1}^{T-l}+$$
$$\sum_{k=i+1}^{p}\ \sum_{l=0}^{j}\ \sum_{s=1-i}^{T-k-i}\ \sum_{t=1}^{T-j}+\sum_{k=i+1}^{p}\ \sum_{l=j+1}^{q}\ \sum_{s=1-i}^{T-k-i}\ \sum_{t=1}^{T-l}\Big\}$$
$$(g_{s+k}f_{t+l}H[s+i,t+j]+g_{s+k}f_{t+j}H[s+i,t+l]+$$
$$g_{s+i}f_{t+l}H[s+k,t+j]+g_{s+i}f_{t+j}H[s+k,t+l])\vartheta_k\alpha_l.\ \square$$

As before we also can form the vectors

$$\frac{\partial(MM^{T})}{\partial\alpha_i}f=\sum_{k=0}^{i}\mathcal{F}(i,k)\alpha_k+\sum_{k=i+1}^{q}\mathcal{F}(k,i)\alpha_k,\ \text{with}$$
$$\mathcal{F}(i,k)=(\underbrace{0\ldots0}_{\leftarrow k\rightarrow}\ f_{i+1}\ldots f_T\ \underbrace{0\ldots0}_{\leftarrow i-k\rightarrow})^{T}+(\ \underbrace{0\ldots0}_{\leftarrow i\rightarrow}\ f_{j+1}\ldots f_{T+j-i})^{T}.$$

and

$$\frac{\partial(P^{T}P)}{\partial\vartheta_j}=\sum_{k=0}^{j}\widetilde{g}(j,k)\vartheta_k+\sum_{k=j+1}^{p}\widetilde{g}(k,j)\vartheta_k,\ \text{with}$$
$$\widetilde{g}(i,k)=(g_{i-k+1}\ldots g_{T-k}\ \underbrace{0\ldots0}_{\leftarrow i\rightarrow})^{T}\ +(\underbrace{0\ldots0}_{\leftarrow i-k\rightarrow}\ g_1\ldots g_{T-i}\ \underbrace{0\ldots0}_{\leftarrow i\rightarrow})^{T}.$$

Corollary 6.3

Second derivatives of AR case

Let $H_4 = X(X^TP^TPX)^{-1}X$. Then

$$\frac{\partial^2 S^*}{\partial\vartheta_i\partial\vartheta_j} = 2e^T\frac{\partial P^T}{\partial\vartheta_i}\frac{\partial P}{\partial\vartheta_j}e - 2e^T\frac{\partial(P^TP)}{\partial\vartheta_i}H_4\frac{\partial(P^TP)}{\partial\vartheta_j}e, \tag{6.16}$$

with

$$e^T\frac{\partial P^T}{\partial\vartheta_i}\frac{\partial P}{\partial\vartheta_j}e = 2\sum_{h=1+\max(i,j)}^{T} e_{h-i}e_{h-j} \tag{6.16.1}$$

and

$$2e^T\frac{\partial(P^TP)}{\partial\vartheta_i}H_1\frac{\partial(P^TP)}{\partial\vartheta_j}e = \{\sum_{k=0}^{i}\sum_{l=0}^{j}\sum_{s=1-k}^{T-i-k}\sum_{t=1-l}^{T-j-l} + \sum_{k=0}^{i}\sum_{l=j+1}^{p}\sum_{s=1-k}^{T-i-k}\sum_{t=1-j}^{T-l-j} + \sum_{k=i+1}^{p}\sum_{l=0}^{j}\sum_{s=1-i}^{T-k-i}\sum_{t=1-l}^{T-j-l} + \sum_{k=i+1}^{p}\sum_{l=j+1}^{p}\sum_{s=1-i}^{T-k-i}\sum_{t=1-j}^{T-l-j}\}$$
$$e_{s+i}e_{t+j}H_1[s+k,t+l] + e_{s+i}e_{t+l}H_1[s+k,t+j] + e_{s+k}e_{t+j}H_1[s+i,t+l] + e_{s+k}e_{t+l}H_1[s+i,t+j])\vartheta_k\vartheta_l. \tag{6.16.2}$$

Proof

Follows directly from (6.13.1) and (6.13.2) as $g = e$ and $H_4 = X(X^TP^TPX)^{-1}X$ as $W^{-1} = (P^{-1}P^{-T})^{-1} = P^TP$.□

Corollary 6.4

Second derivatives of MA case

Let $h = (MM^T)^{-1}e$ and $H_5 = (M^{-T}M^{-1} - M^{-T}M^{-1}X(X^TM^{-T}M^{-1}X)^{-1}X^TM^{-T}M^{-1})$. Then

$$\frac{\partial^2 S^*}{\partial\alpha_i\partial\alpha_j} = -2\, h^T\frac{\partial M}{\partial\alpha_i}\frac{\partial M^T}{\partial\alpha_j}h + 2h^T\frac{\partial(MM^T)}{\partial\alpha_i}H_5\frac{\partial(MM^T)}{\partial\alpha_j}h \tag{6.17}$$

$$h^T\frac{\partial M}{\partial\alpha_i}\frac{\partial M^T}{\partial\alpha_j}h = \sum_{h=1}^{T-\max(i,j)} h_{h+i}h_{h+j} \tag{6.17.1}$$

and

$$h^T\frac{\partial(MM^T)}{\partial\alpha_i}H_5\frac{\partial(MM^T)}{\partial\alpha_j}h=\{\sum_{k=0}^{i}\sum_{l=0}^{j}\sum_{s=1}^{T-i}\sum_{t=1}^{T-j}+\sum_{k=0}^{i}\sum_{l=j+1}^{q}\sum_{s=1}^{T-i}\sum_{t=1}^{T-l}+\sum_{k=i+1}^{q}\sum_{l=0}^{j}\sum_{s=1}^{T-k}\sum_{t=1}^{T-j}+$$
$$\sum_{k=i+1}^{q}\sum_{l=j+1}^{q}\sum_{s=1}^{T-k}\sum_{t=1}^{T-l}\}(h_{s+k}h_{t+l}H_5[s+i,t+j]+h_{s+k}h_{t+j}H_5[s+i,t+l]+$$
$$h_{s+i}h_{t+l}H_5[s+k,t+j]+h_{s+i}h_{t+j}H_5[s+k,t+l])\alpha_k\alpha_l. \quad (6.17.2)$$

Proof

Obvious from (6.14), (16.14.1) and (16.14.2) as P=I. □

6.6 Conclusion

In this chapter the well known Conditional Least Squares function is treated. Both first and second derivatives are given. The Yule-Walker conditions are a special case, treated in the first corollary. The derivations are tedious and lengthy but give no special problems. Remarkably is that the expressions are still complicated, even more than in the exact case. Reason of this phenomenon is that the covariance matrix has no longer diagonals of which all elements are equal. Of course the number of elements of the derivatives are less than in the exact case as the determinant part is absent.

VII SIMULATIONS

7.1 Maximum Likelihood and Conditional Least Squares

The method of maximum likelihood (ML) to estimate ARMA parameters is since long widely recommended. The main reason is the fact that maximum likelihood estimators possess nice asymptotic properties, a second reason is that simulation evidence suggests better performance over alternatives (see *e.g.* Ansley and Newbold, 1980). The ML method has one severe problem, as both inverse and determinant of the covariance matrix need to be available. Without a closed form for the covariance matrix the computation of both is an time consuming task.
Alternatives to ML estimation are Minimum Distance estimation and Conditional Least Squares estimation, the most used method to estimate ARMA parameters. As argued in the preceding chapter we will use CLS to compare with ML. It differs from ML estimation as the determinant of the covariance matrix to the power of reciprocal of the number of observations is disregarded and as the covariance matrix is replaced by an approximation.

Between these two methods, ML and CLS, one finds a large number of methods that in one way or another avoid the problem of putting the starting values to zero. We will mention a few of them. One of the oldest is the well known back forecasting (sometimes called back casting) method of Box and Jenkins (1976, p. 200), which is at the same time notorious for its bad computational properties. Kohn and Ansley (1985) presented formulas for the likelihood and its derivatives. Their approach is based upon replacing the ARMA(p,q) process by a MA(q) process for the first T-p (!) observations. In short, all methods proposed give an alternative to avoid the zero starting values. More recently, several authors, like Zinde-Walsh (1988) and Knottnerus (1989), who uses a Kalman-filter, claim to have developed algorithms to find the exact covariance matrix. Their methods, like most others are very complicated. As all methods are meant as an approximation for the exact likelihood we will not further investigate them.

Thus we will concentrate upon the differences between (exact) ML and CLS, as they can be seen as each others opposites. In general, the differences between both are small, but the overall performance of the ML estimates seems to be slightly better.

7.2 An Algorithm to compute ARMA parameters

Before we turn to our experiments we give an outline of the algorithm we use. Equivalent to maximising the likelihood function is minimising the function $S(\alpha,\vartheta) = |V|^{1/T} e^T V^{-1} e$, where V, apart of a constant, is the ARMA covariance matrix and $e = y - Xb$, where $b = (X^T V^{-1} X)^{-1} X^T V^{-1} y$, the Aitken estimator of β. Starting with the unit matrix for V, b becomes the OLS estimate $(X^T X)^{-1} X^T y$; the residuals are computed as y-Xb. The estimated values for α and ϑ are found where $S(\alpha,\vartheta)$ attains its minimum. These values, say α_0 and ϑ_0 are used to find a new estimate for β, using $V_0 = V(\alpha_0,\vartheta_0)$, which gives new values for e. The process stops when the value of b (and thus α and ϑ) does not change more than a certain threshold value.
Of course the parameter values have to obey the constraints imposed by the invertibility conditions. Unfortunately, these are in implicit form, which makes it difficult to employ standard software for minimisation problems. That is the reason why we developed a new algorithm, using MATLAB.

Of course the problem is to find an efficient way to compute new values for α and ϑ. Our approach is Newton-like and we proceed as follows. First compute **h**, the vector of first derivatives and **H**, the (Hessian) matrix of second derivatives of the modified likelihood function $S(\alpha,\vartheta)$. Next find a new point for the parameter vector by computing the point where the tangent dissects the parameter axis: $-\mathbf{H}^{-1}\mathbf{h}$. Again the first derivative is computed, and a new dissection point is found. In this way the zeros of the first derivative are found, but these need of course not to correspond to a minimum. Therefore we modify the Hessian matrix.
The eigenvalues of the Hessian matrix tell us that we are in the neighbourhood of a minimum if all eigenvalues are positive, or equivalently, that the function is locally convex. If the eigenvalues are not positive we may find a maximum or a saddle point. To avoid this situation we change the sign of the negative eigenvalues and compute a modified Hessian, say

$\tilde{H}$, which is positive definite. This assures us that we will at least not search in the direction of a maximum. This method was proposed by Greenstadt (1967). According to Kennedy and Gentle (1980, p.443) it seems to work well in practice.
Naturally there are more pitfalls, *e.g.* when the modified Hessian matrix is almost singular, giving very large values of $a = -\tilde{H}^{-1}\mathbf{h}$. Therefore we divide a by $\max(1, a^T a)$, giving $\tilde{a}$. But even then, there is no guarantee that the new parameter value is within the constraints imposed by the invertibility conditions. Denoting the new and old parameter values by ξ_1 and ξ_0 we chose $\xi_1 = \xi_0 + \lambda \tilde{a}$, where λ is such that the value of ξ_1 is admissible in view of the invertibility condition and the corresponding function value $S(\xi_1)$ is smaller than $S(\xi_0)$. At the same time this procedure prevents the algorithm from switching between two points.
The Hessian matrix is not recomputed at every step. Recomputation depends on the number of positive eigenvalues and the speed of decrease of the derivatives in absolute value. As starting point for the ML estimates the results of the CLS method are used, which can be regarded as an approximation of the maximum likelihood estimates.
The process stops if one of the following quantities is smaller than the threshold (which we fixed at 10^{-6}):

1. the maximum of the absolute value of the derivatives, indicating, that a stationary point is reached;
2 the maximum of the difference between two successive derivatives, indicating a stationary point or a boundary point;
3 the absolute difference of two successive function values;
4 the maximum of the absolute change in parameter values.

Experiments show, that these latter criteria have to be used, as in several cases a boundary is reached before a minimum. This means, that the findings of Cryer and Ledolter (1981), who investigated the MA(1) case only, are not generally valid.

Of course such an algorithm only brings us to a local minimum, if it exists. Therefore we performed several experiments with different starting values, which -as far as our experience reaches- always gave the same minimum. However, when the number of parameters is moderate or large, say 5 or more, there is no simple way to detect more or other local minima. In many cases not the abundance of minima gives the problem, but the flatness of the likelihood function, giving a boundary point.

7.3 Technical description of the algorithm

The experiments were carried out as follows. Values were chosen for T, p, q, $\vartheta_i, i=1,..,p$, $\alpha_i, i=1,..,q$, $\beta(=1)$ and eps $(=10^{-6})$. Next errors were generated by $\varepsilon_t = \sum_{h=1}^{p} \vartheta_h \varepsilon_{t-h} + \sum_{h=0}^{q} \alpha_h u_{t-h}$, $t=1,..,T$, where u_t is white noise, and $y_t=\beta+\varepsilon_t$, $t=1,..,T$ was computed. The algorithm, which will be described below, was used to find estimates for the parameter vector and β. Apart of the symbols already used we define $r=\max(p,q)$, $f=|V|^{1/T}e^TV^{-1}e$ and $\xi^T=[\vartheta^T\ \alpha^T]$. As starting values were chosen for $\xi^0(=\xi)$, $b^0(=1)$, $b(=0)$.

The algorithm consists of the following steps:
1 If $(b-b^0)>$eps stop.
1.1 Compute e and b, using (2.7):

1.1.1 $Z=[X\ y] \in \mathbb{R}^{Tx(k+1)}$, then $Z^TV^{-1}Z=\begin{bmatrix} X^TV^{-1}X & X^TV^{-1}y \\ y^TV^{-1}X & y^TV^{-1}X \end{bmatrix}$.

Compute $Z^TV^{-1}Z=Z^TP^TM^{-1}\{I_T-R(R^TR+\underline{D})^{-1}R^T\}M^{-1}PZ \in \mathbb{R}^{(k+1)x(k+1)}$

1.1.2 Use *MinvP* to find $\gamma_i, i=1,..,T^*$, the elements of the lower band matrix $M^{-1}P$ and its length T^*.

1.1.3 Form $Z^*_{i,j}(=M^{-1}PZ) = \sum_{h=1}^{\min(T^*,i)} \gamma_i Z_{i-h+1,j} \in \mathbb{R}^{T^*x(k+1)}$.

1.1.4 $R=M^{-1}PN-Q \in \mathbb{R}^{T^*xr}$.
1.1.5 $H_1=R^TZ^* \in \mathbb{R}^{rx(k+1)}$.
1.1.6 Use *DELTA* to find $\underline{D}=\underline{P}^T\underline{P}-\underline{QQ}^T \in \mathbb{R}^{rxr}$.
1.1.7 $H_2=(R^TR+\underline{D})^{-1} \in \mathbb{R}^{rxr}$.
1.1.8 $Z^TV^{-1}Z=Z^{*T}Z^*-H_1^TH_2H_1$.
1.1.9 $b=(X^TV^{-1}X)^{-1}X^TV^{-1}y$ and $e=y-Xb$.

1.2 Compute f, using *FVAL*.
1.3 Put $v=$ 2eps, $\lambda=1$, condition$=-1$.
1.4 If condition>0 go to 2.

1.5 Compute df, the vector of derivatives of f to ξ and H. H is a direction matrix which is found as follows:
- compute the matrix of second derivatives
- make it positive definite by changing the sign of negative eigenvalues
- invert the resulting matrix.

1.6 condition = # positive eigenvalues.

2 If $v<eps$ go to 1. (Given b a minimum of $f(\xi)$ has been found.)
2.1 Put $\xi^0=\xi$, $f^0=f$, $df^0=df$.
2.2 Compute new values for ξ, f, df, H, condition, and λ.
2.2.1 Find new direction: $a=-H\ df^0$.
2.2.2 $c=\max(1,a^Ta)$.
2.2.3 $aa=\max|a_i/c|$, $i=1,..,p+q$.
2.2.4 $f=2f^0$, $\lambda=\min(1.25*\lambda,1)$, $\xi=\xi^0$.
2.2.5 If $f\le f^0$ or $aa*\lambda\le eps$ go to 2.2.9.
2.2.5.1 $k=1$.
2.2.5.2 If $k\le 0$ or $aa*\lambda\le eps$ go to 2.2.6.
2.2.5.3 $\xi=\xi^0+\lambda*a$.
2.2.5.4 If ξ is admissible put $k=0$.
2.2.5.5 Go to 2.2.5.2.
2.2.6 Compute f, using *FVAL*.
2.2.7 $\lambda=0.9*\lambda$.
2.2.8 Go to 2.2.5.
2.2.9 Compute new derivatives df.
2.2.10 If $df^{0^T}df^0<df^Tdf$ then condition=condition-2.5 else condition=condition+.5.
2.2.11 If condition<0 then
compute new H-matrix (see 1.5)
condition = #positive eigenvalues
$\lambda=1$.
2.3 $v_1=\max|df_i|$, $i=1,..,p+q$,
$v_2=\max|df_i-df_i^0|$, $i=1,..p+q$,
$v_3=|f-f^0|$,
$v_4=\max|\xi-\xi^0|$.
$v=\max(v_1,v_2,v_3,v_4)$.
2.4 Go to 2. ■

FVAL

FVAL is used to compute the concentrated likelihood, $|V|^{1/T}e^TV^{-1}e$.

1 Compute $|V|$, using expression (2.8).

1.1 Use *LBM* to find $\underline{P}$ and $\underline{M} \in \mathbb{R}^{rxr}$.

1.2 Use *UBM* to find $\underline{Q}$ and $\underline{N} \in \mathbb{R}^{rxr}$.

1.3 $H=\underline{PN}\text{-}\underline{MQ}$.

1.4 Use *DELTA* to find $\underline{D}=\underline{P}^T\underline{P}\text{-}\underline{QQ}^T \in \mathbb{R}^{rxr}$.

1.5 Construct $M^* \in \mathbb{R}^{T^*xr}$, the first r columns of M, using *INVEC.*

1.6 $|V| = |I_r+\underline{D}^{-1}H^TM^{*T}M^*H|$.

2 Compute $e^TV^{-1}e$, using expression (2.7).

2.1 Use *MinvP* to find γ_i, $i=1,..,T^*$, the elements of the lower band matrix $M^{-1}P$ and its length T^*.

2.2 Compute $e_i^* = \sum_{h=1}^{i} \gamma_{i-h+1}e_h$, $1 \le i \le T^*$ and $e_i^* = \sum_{h=1}^{T^*} \gamma_{T^*-h+1}e_{i-T^*+h}$, $T^*+1 \le i \le T$.

2.3 $R=M^{-1}PN\text{-}Q \in \mathbb{R}^{Txr}$.

2.4 $h=R^Te^*$.

2.5 $H=(R^TR+\underline{D})^{-1} \in \mathbb{R}^{rxr}$.

2.6 $e^TV^{-1}e=e^{*T}e^*\text{-}h^THh$.

3 $FVAL=|V|^{1/T}e^TV^{-1}e$. ■

LBM

LBM gives the lower band matrix $Z \in \mathbb{R}^{N*N}$ using the vector $\xi=(\xi_0\ \xi_1..\xi_n)^T$:

1 For $j=1$ to $N\text{-}n$; for $i=j$ to $j+n$; $Z(i,j)=\xi_{i-1}$.

2 For $j=N\text{-}n+1$ to N; for $i=1$ to $\min(N\text{-}j,n)$; $Z(i,j)=\xi_{i-1}$. ■

UBM

UBM gives the upper band matrix $Z \in \mathbb{R}^{NxN}$, using the vector $\xi=(\xi_0\ \xi_1..\xi_n)^T$:

1 For $j=N\text{-}n+1$ to T ; for $i=1$ to $n+j\text{-}N$; $Z(i,j)=\xi_{N+i-j}$. ■

INVEC

Given $\xi_i, i=0,..,n$, the elements of a lower band matrix of dimension $N{\times}N$, *INVEC* gives x_i, $i=1,..N^*$, the N^* (non-zero) elements of the lower band matrix that is the inverse of the former.

1 $x_1=1/\xi_1$.

2 For $i=1$ to N; $x_i = - \sum_{h=\max(1,i-p)}^{i-1} x_1 x_h \xi_{i-h}$.

3 Find N^* such that for all $i \geq N^*$ $\xi_i < 10^{-10}$. ■

DELTA

DELTA gives $Z \in \mathbb{R}^{nxn}$ (the inverse of the AR covariance matrix) using $\xi = (\xi_0\ \xi_1..\xi_n)^T$.

1 $Z = \sum_{h=0}^{n-1} (\xi_h \xi_{i-j+h} - \xi_{i+h} \xi_{j+h})$. ■

MinvP

MinvP computes γ_i, $i=1,..,N^*$, the (non-zero) elements of the (NxN) lower band matrix $M^{-1}P$ and its length N^*, using α_i, $i=1,..,q$, the elements of M and ϑ_i, $i=1,..,p$, the elements of P.

1 Use *INVEC* to compute a and N^* from α, q and N.

2 For $i=1$ to $\min(N^*, p+q+1)$; $\gamma_i = \sum_{\max(i,0)}^{\min(p,i-N^*)} a_{i-h} \vartheta_h$. ■

7.4 The experiments

Two types of experiments were carried out. The former consisted of experiments estimating various ARMA models. The latter series were repeated experiments of the same models with different random numbers. We will discuss both separately.

For first type of the experiments we employed twenty-four (5^2-1) different ARMA models, *viz.* ARMA(p,q), p=0,..4, q=0,..,4, with the obvious exception p=q=0. For every model we simulated time series with successively 20, 40, 60, 75, 150 and 300 observations. Furthermore the experiments were carried out three times, using different seeds for the (MATLAB) random generator. Time series were generated using normally distributed random numbers and with the number of parameters to be estimated. To keep things simple we have chosen $\beta=1$ and a vector of ones for X.

We choose more or less arbitrary values for the ARMA parameters. The only restriction is that they had to satisfy the constraints of the invertibility conditions and that the associated polynomials of the AR and MA parts had no common zeros.

parameter values					corresponding roots			
AR	.8	.3	-.2	.1	-0.64-0.61i	-0.64+0.61i	0.24-0.27i	0.24-0.27i
MA	-.5	.6	.2	-.1	0.32-0.84i	0.32+0.84i	-0.43	0.28

Any time two estimates were made, one using the conditional least squares method, the other one using the exact likelihood function. Every run consists of 6 series of (5*5-1) parameters for both methods, giving 124 simulation runs for both methods. These simulations were done three times, using different seeds for the random generator, which brings the total number of simulation runs for each method on 432.

Questions we are interested in are:

1. How fast is the algorithm?
2. Which method is 'best' with regard to the estimation of the regression coefficient?
3. Which method is 'best' with regard to the estimation of ARMA parameters?

1. Speed

The algorithm works quite fast. A minimum is found in a few steps, surely in the case of a pure MA or AR case. This minimum can be attained at a boundary point. Usually the initial estimate of the parameters takes most time, consecutive estimates take mostly less than three steps. In several cases two different starting points for the parameter vector are chosen, *viz.* the zero vector and the value of the vector with which the time series were generated. In all cases they gave equal results.

2. Regression coefficient estimation

For an evaluation of the estimate of the regression coefficient we counted the number of times that the ML estimate was closer to the true value than the CLS estimate. The results are summarized in Table 1.

Table 1.

T \ Seed	0	500	1000	Total
20	6	4	0	10
40	0	2	-4	-2
60	4	4	-4	4
75	4	4	-2	6
150	4	2	-2	4
300	-8	2	2	-4
Total	10	18	-10	18

Number of times ML performs better than CLS in estimating β

In Table 1. one sees which method performs better in estimating the regression coefficient. The first column (T) gives the number of observations, column 2, 3 and 4 give the results for the different seeds used for the random generator. Every ML estimate better than the CLS estimates counts for one, worse for minus one. Every cell is based on 24 experiments, *viz.* all ARMA(p,q) models from 0,0 to 4,4 minus the (0,0) case. The overall sum (18 out of 3*6*24) indicates a slightly better result for the ML method. Remarkably is the role of the stochastic component. While in the second case (with seed= 500) ML performs always better than CLS, the opposite is almost true in the last case (with seed= 1000), where CLS performs clearly better.

3. Parameter estimates

Table 2 to gives an idea of the quality of the estimates of the parameter values. As criterion is used the sum of squared differences between the estimates and the true parameter values.

Table 2.

T	0	500	1000	Total
20	0	-1	-4	-5
40	-5	-6	1	-10
60	-5	-1	7	1
75	-2	6	10	14
150	3	0	3	6
300	4	6	-2	8
Total	-5	4	15	14

Number of times ML performs better than CLS in estimating ARMA-parameters.

Again, the first column (T) gives the number of observations, column 2, 3 and 4 give the results for the different seeds (0, 500, 1000) used for the random generator. When ML performs better, it is counted for one else for minus one. The overall score is positive, indicating a better performance for the ML approach, but striking is the bad result for short time series. A possible cause can be the preference of ML estimates to give boundary values. We will have a closer look at it in the sequel.

Table 3.

	AR		MA		Total
T	CLS	ML	CLS	ML	
20	8	6	18	31	63
40	0	0	16	31	47
60	1	1	9	23	34
75	0	0	5	18	23
150	1	1	6	6	14
300	0	0	0	0	0
Total	10	8	54	109	181

Boundary values.

In Table 3 gives a summary of the number of times a unit root is found. As criterion we employ an (absolute) value greater than .999. Every cell regards 3 times 24 experiments: the results for the three different values of the seed are taken together. The conclusions are obvious:

- the smaller the number of observations, the greater the number of unit roots
- the estimation of MA parameters results in more unit roots than the estimation of AR parameters

- ML estimation of MA parameters results in more unit roots than CLS estimation.

It is well known that the likelihood function can be very flat, especially when the number of observations is small and individual observation can have a strong influence on the function. We refer to the article of Cryer and Ledolter (1981). When one tries to estimate 5 or more ARMA parameters using only 20 observations, one can expect that a boundary value will be found quite often.

In most cases the derivative to the MA parameter in the ML case is zero or close to zero, indicating a minimum of the likelihood function on the boundary, but this need not always be the case: we did find experiments with unit roots while the derivatives were not close to zero. A unit root for the AR parameters should be rare as the inverse of the covariance matrix becomes singular at the boundary. This implies that the determinant of the covariance matrix will become very large: the boundary can never be reached exactly. Indeed, we rarely found this situation, mostly when the number of observation was small ($T=20$) and the number of parameters to estimate large (5 or more). The sensitivity of ML estimation compared to CLS for unit roots is not new. It is *e.g.* reported by P. Newbold *et al.* (1994).

We performed another type of simulations hoping to be able to give a more decisive opinion about the behavior of both methods and to get insight into the variances of the estimators. For this end we fixed the number of simulations at 2500. The first experiment concerned an ARMA(2,2) model with parameters $\vartheta=(.8\ \ .3)$ and $\alpha=(-.5\ \ .6)$. The roots are -4.±.37i and .25±.73i. The results are summarized in Table 4.

Table 4.

	CLS		ML	
N	2453		2466	
$\hat{\beta}-\beta$	-.0022	(.007)	-.0018	(.007)
$\hat{\vartheta}_1-\vartheta_1$	.0435	(.111)	.0372	(.106)
$\hat{\vartheta}_2-\vartheta_2$	.0472	(.077)	.0441	(.075)
$\hat{\alpha}_1-\alpha_1$	.0764	(.127)	.0215	(.127)
$\hat{\alpha}_2-\alpha_2$	.0814	(.075)	.0271	(.096)

Estimation results for an ARMA(2,2) model with $T=40$, $\vartheta=(.8\ .3)$, $\alpha=(-.5\ .6)$.

The first row, N, gives the number of simulations (out of 2500) where an internal point was found, in the remaining cases a boundary point resulted. Between parentheses one reads the variance of the corresponding estimate. The overall view is a slightly better performance of ML over CLS, but the differences are small. Remarkable is that N is higher for the ML method.

Similar simulations were performed for two ARMA(2,1) models with different parameter values. In both cases the number of observations was 30, the number of simulations again 2500. The models are $(\vartheta, \alpha)=(.8, .3, -.5)$ with roots $.4\pm.37i$ and $.5$ and $(\vartheta, \alpha)=(.4, .2, .25)$ with roots $.2\pm.4i$ and $-.25$. The differences between both models are dramatic.

Table 5.

	CLS		ML	
N	1983		1254	
$\hat{\beta}-\beta$	.0006	(.002)	-.0006	(.002)
$\hat{\vartheta}_1-\vartheta_1$	.0087	(.091)	.1105	(.080)
$\hat{\vartheta}_2-\vartheta_2$	.0410	(.064)	.1291	(.060)
$\hat{\alpha}-\alpha$	.0705	(.124)	.1326	(.115)

Estimation results for ARMA(2,1) with $T=30$, $\vartheta=(.8\ .3)$, $\alpha=-.5$

Table 6.

	CLS		ML	
N	1884		1659	
$\hat{\beta}-\beta$	.0001	(.021)	.0001	(.021)
$\hat{\vartheta}_1-\vartheta_1$	.0054	(.227)	-.0045	(.194)
$\hat{\vartheta}_2-\vartheta_2$	.0700	(.042)	.0677	(.044)
$\hat{\alpha}-\alpha$	-.0117	(.244)	.0193	(.200)

Estimation results for ARMA(2,1) with $T=30$, $\vartheta=(.4\ .2)$, $\alpha=.25$

While Table 6 gives slightly better results for ML over CLS, Table 5 gives the opposite view. Moreover in almost half of all simulations ML was not able to find an estimate within the admissible region.

7.5 Summary

In this chapter we presented an algorithm to compute the ARMA parameters from the data for a regression model with ARMA distributed errors. Of course the same algorithm can be used to compute the parameters of an exact time series model. The only difference is the computation of the error vector in case of a regression model. The error vector depends on the ARMA parameters used and an iterative procedure has to be used.

The algorithm we use is extensively described, including efficient ways to design the basic building blocks of our approach, *viz.* the lower and upper band matrices, their inverses and several products of these matrices, including the covariance matrix itself. Although the concentrated likelihood function contains the determinant and the inverse of a TxT matrix, where T stands for the number of observations, we present algorithms where only matrices of order max(p,q), the higher number of AR and MA parameters have to be inverted. The computation of the determinant of the TxT covariance matrix can be reduced to the computation of, again, a matrix of order max(p,q).

The algorithm to compute estimates of the ARMA parameters works fast. The optimum of the function of interest, within the admissible region, is mostly found within a few steps. Moreover we found - as far as our experience reaches - the same optimal point regardless of the starting values. An reason for this phenomenon can be the well known flatness of the likelihood function when the number of dimensions is more than one or two. The algorithm is used to find an answer to the much debated question whether ML estimation performs better than CLS estimation. Although ML is preferred because of its nice theoretical properties it involved a high penalty over CLS thus far because of its computational burden. Moreover it has not always been clear whether the estimation procedure used was exact maximum likelihood or an approximation in one way or another, like the backforcasting method of Box and Jenkins.

The answer however is not clear-cut. In total we performed 5432 simulation runs. The performance of exact maximum likelihood is slightly better than CLS, but the differences are small. The results of the simulations depend strongly upon the random numbers used, which makes a definite answer difficult. To conclude, we recommend exact ML over CLS for two reasons: first exact ML hardly takes more computations using our techniques than CLS, secondly it is to be preferred on theo-retical grounds as CLS is only an approximation for ML.

SUMMARY

This study presents a closed form expression for the covariance matrix of consecutive observations of an ARMA process and investigates its applications for exact maximum likelihood estimation of models with ARMA errors.

ARMA errors are often applied in pure time series models. Its use in regression models is mostly restricted to an AR(1) or MA(1) process. Such a distinction between time series and regression models is not necessary. It is easy to show that in both cases the same weighted sum of squares as function of the ARMA parameters is minimised when minimum distance estimators are used. Maximum Likelihood estimation (ML) under the usual assumptions is in both cases equivalent to minimisation of the same sum of squares times a function of the determinant of the covariance matrix, for which we use the name modified likelihood function. The difference between both methods is the way the observations of an ARMA process are obtained: in a pure time series model they are given, in a regression model they are estimated as the residuals, the difference between observed and computed values of the dependent variable. As these residuals depend on the unknown ARMA parameters an iterative procedure has to be employed.

For efficient computation of this sum of squares and the corresponding determinant a closed matrix form of the covariance matrix of the ARMA errors is required. Until now only very time consuming and complicated algorithms were available to compute the individual elements of the covariance matrix. Only for AR(1) and AR(2) and the general MA case such an expression was known.

One of the main results of this study is the covariance matrix in closed form for the general ARMA case. Starting point is the presentation of the ARMA model in matrix form. For both the MA and AR parameters two lower band matrices are formed that describe the ARMA error structure. From these matrices a matrix equation with the covariance matrix as unknown variable is derived. Its solution is trivial in case of a pure MA model, it

is less simple for a pure AR case and, to be honest, it took a lot of effort to find the general solution. The well known invertibility and stationarity conditions can be formulated as conditions of the eigenvalues of a matrix of which the dimension is equal to the number of parameters. The inverse and the determinant of the ARMA covariance matrix are also expressions of these band matrices, where only a matrix with dimension equal to the higher number of AR or MA parameters needs to be inverted. The pure MA and AR case are special cases of the general form. Their covariance matrix is of course a more simple expression than the general covariance matrix.

The covariance matrix itself is simple enough to be differentiated. To achieve this we express the band matrices as a function of the parameters and a lag matrix. A lag matrix is defined as a zero-one matrix with the ones along one of its (off-) diagonals. In this form the band matrices and the covariance matrix can easily be differentiated. This enables us to derive analytical expressions for the first and second derivatives of the likelihood function, which can be used to find estimates for the parameters. The first derivatives look like linear functions of the parameters of interest, but the coefficients of this function strongly depend on the parameters itself. The elements of this matrix of coefficients are the sum of two parts. The first one corresponds to the quadratic part and is quadratic itself, the second part stems from the determinant part and consists of a function of individual elements of the covariance matrix. In the pure MA or AR case they reduce to the sum of elements of the main or off diagonals. The second derivative consists of five, relatively simple, parts. One of these is the inverse of the information matrix.
Until now one of the most common methods to estimate ARMA parameters is the so called Conditional Least Squares method, shortly CLS. This method circumvents the problem of the lack of the exact covariance matrix by disregarding starting values. Hence its name, as its results are conditional on the starting values. Using the exact covariance matrix we can show that the sum of squares to be minimised are asymptotically equal in the exact ML and the CLS case. The determinant however is not: its value is always equal to one in the CLS case and is a function of the ARMA parameters in the exact case, bounded with respect to the number of observations. However for maximum likelihood estimation one needs the value of the determinant to the power one divided by the number of observations, which goes

rapidly to one when the number of observations increases. Hence CLS and exact ML are asymptotically equivalent. As is done for the ML approach we also present first and second derivatives for the CLS function. While the number of elements is less than those in the exact case, they are themselves more complicated. Reason is the fact that the symmetry of the covariance matrix is now less than in the exact case.

We conclude by presenting an algorithm to estimate the ARMA parameters. The algorithm traces the zeros of derivative of the modified likelihood function by a Newton like approach. This algorithm, which takes also invertibility and stationarity restrictions into account, works quite fast. The differences between exact maximum likelihood estimation and conditional least squares estimation are small. We conclude that exact estimation is slightly better. As it takes hardly more computations we recommend the exact way, both on theoretical and on practical grounds.

SAMENVATTING

Deze studie geeft een gesloten vorm uitdrukking voor de covariantie matrix van opeenvolgende waarnemingen van een ARMA proces en onderzoekt de toepassingen er van voor het schatten van modellen met ARMA verstoringen met behulp van de methode van maximale aannemelijkheid.

ARMA verstoringen worden vaak toegepast in zuivere tijdreeks modellen. Het gebruik er van in regressie modellen is meestal beperkt tot een AR(1) of MA(1) proces. Een dergelijk onderscheid tussen tijdreeks en regressie modellen is niet nodig. Het valt gemakkelijk aan te tonen, dat in beide gevallen dezelfde gewogen kwadratensom als functie van de ARMA parameters wordt geminimaliseerd in geval minimum afstand schatters worden gebruikt. Schatten met behulp van maximale aannemelijkheid (verder aan te duiden met de engelstalige afkorting van *maximum likelihood*, ML), onder de gangbare vooronderstellingen is in beide gevallen equivalent met het minimaliseren van dezelfde kwadratensom vermenigvuldigd met een functie van de determinant van de covariantiematrix, de z.g. gemodificeerde aannemelijkheidsfunctie. Het verschil tussen beide methoden is de manier waarop de waarnemingen verkregen worden: in een tijdreeksmodel zijn deze gegeven, in een regressiemodel zijn het de geschatte fouten, het verschil tussen waargenomen en berekende waarde van de afhankelijke variabele. Omdat de berekende waarde afhangt van de onbekende ARMA parameters moet een iteratieve werkwijze worden gevolgd.

Voor het efficiënt berekenen van deze kwadratensom en de bijbehorende determinant is een gesloten matrix vorm voor de covariantie matrix van de ARMA verdeelde verstoringen vereist. Tot nu toe waren er alleen zeer tijdrovende en ingewikkelde algoritmes beschikbaar om afzonderlijke elementen van de covariantie matrix te berekenen. Allen voor AR(1) en AR(2) en het algemene MA-geval was een dergelijke uitdrukking voor handen.

Een van de belangrijkste resultaten van deze studie is de gesloten matrix vorm voor de covariantiematrix voor het algemene ARMA geval. Vertrekpunt

is de presentatie van de AR en MA parameters in de vorm van beneden band matrices. Voor zowel de MA als AR parameters worden twee beneden band matrices gevormd die de ARMA verstoringenstructuur beschrijven. Met behulp van deze matrices wordt een matrix vergelijking met de covariantie matrix als onbekende grootheid afgeleid. De oplossing er van is triviaal in het geval van een puur MA model, iets minder eenvoudig voor het pure AR model en - om eerlijk te zijn - het kostte de nodige inspanning om de algemene oplossing te vinden. De bekende voorwaarden ten aanzien van inverteerbaarheid en stationariteit worden geformuleerd als eigenwaarden condities van een matrix waarvan de dimensies gelijk zijn aan het aantal parameters. De inverse en de determinant van de ARMA covariantie matrix zijn eveneens uitdrukkingen van de band matrices, waarbij slechts een matrix met dimensie gelijk aan het hoogste aantal van de MA of AR parameters hoeft te worden geïnverteerd. De zuivere MA en AR gevallen zijn speciale gevallen van de algemene vorm en hebben een eenvoudiger uitdrukking dan de algemene covariantie matrix.

De covariantie matrix zelf is eenvoudig genoeg om analytisch gedifferentieerd te worden. Om dit te bereiken schrijven we de band matrices als een functie van de parameters en een vertragingsmatrix. Een vertragingsmatrix is gedefinieerd als een nul-een matrix met enen op een van zijn (zij-)diagonalen. In deze vorm kunnen band matrices en de covariantie matrix eenvoudig worden gedifferentieerd. Dit stelt ons in staat analytische uitdrukkingen af te leiden voor de eerste en tweede afgeleiden van de aannemelijkheidsfunctie, die gebruikt kunnen worden bij het zoeken naar schattingen voor de parameters. De eerste afgeleiden zien er uit als lineaire functies van de desbetreffende parameters, maar de coëfficiënten van deze functie hangen sterk af van de parameters zelf. De elementen van deze coëfficienten matrix zijn de som van twee delen. Het eerste deel correspondeert met het kwadratische deel en is ook weer kwadratisch, het tweede deel is afkomtig van het determinant gedeelte en bestaat uit een functie van de afzonderlijke elementen van de covariantie matrix. In de zuivere MA en AR gevallen reduceert dit tot de som van elementen van de hoofd- of nevendiagonalen. De tweede afgeleide bestaat uit vijf, op zich eenvoudige delen, waarvan er een de inverse van de informatiematrix is.

Tot op heden is de z.g. Conditionele Kleinste Kwadraten methode (die we aanduiden met de engelstalige afkorting van *Conditional Least Squares*, CLS) een van de meest gebruikte methoden om ARMA parameters te schatten. Deze methode omzeilt het probleem van het gemis van de exacte covariantie matrix door de startwaarden te verontachtzamen. Vandaar de naam, omdat de resultaten voorwaardelijk zijn ten aanzien van de startwaarden. Door gebruik te maken van de exacte covariantie matrix kunnen we laten zien, dat de te minimaliseren kwadratensom asymptotisch gelijk in geval van exacte ML en het CLS geval. Dat geldt niet voor de determinant: die heeft altijd de waarde een in het CLS geval en is een ten opzichte van het aantal waarnemingen begrensde functie van de ARMA parameters in het exacte ML geval. Voor ML schatting telt evenwel de waarde van de determinant tot een macht gelijk aan de reciproke van het aantal waarnemingen, iets dat snel tot een nadert wanneer het aantal waarnemingen toeneemt. Vandaar dat CLS en ML asymptotisch equivalent zijn. Zoals gedaan is voor de ML benadering presenteren we ook eerste en tweede afgeleiden voor de CLS functie. Terwijl het aantal elementen kleiner is dan in het exacte geval, zijn ze op zich ingewikkelder. Reden daarvoor is dat de symmetrie nu kleiner is dan in het exacte geval.

We besluiten met het geven van een algoritme om de ARMA parameters te schatten. Het algoritme spoort de nulpunten op van de eerste afgeleide met een Newton-achtige benadering. Dit algoritme, dat rekening houdt met de inverteerbaarheids- en stationariteitsvoorwaarden, werkt tamelijk snel. De verschillen tussen exacte maximale aannemelijkheid en conditionele kleinste kwadraten zijn gering. We concluderen dat exacte schatting lichtelijk beter werkt. Omdat er nauwelijks meer berekeningen nodig zijn bevelen we de exacte methode aan, zowel om theoretische als praktische redenen.

REFERENCES

Anderson, T.W., 1971, The Statistical Analysis of Time Series, Wiley, New York.

Anderson, T.W. and R.P. Mentz, 1982, Maximum likelihood estimation in autoregressive and moving average models, in Time Series Analysis: Theory and practice 1, O.D. Anderson (ed).

Ansley, C.F. and P. Newbold, 1980, Finite sample properties of estimates for autoregressive moving average models, Journal of Econometrics, 13, 159-183.

Beach, C.M. and J.G. MacKinnon, 1978, A Maximum Likelihood Procedure for Regression with Autocorrelated Errors, Econometrica, 46, 51-57.

Box, G.E.P. and G.M. Jenkins, 1976, Time Series Analysis, Holden Day Inc., San Francisco.

Cryer, J.D. and J. Ledolter, 1981, Small sample properties of the maximum likelihood estimator in the first order moving average model, Biometrika, 68, 691-694.

Davies, J.Ph., 1979, Circulant Matrices, Wiley, New York.

Diebold, F.X., 1986, The exact initial matrix of the state vector of a general MA(q) process, Economics Letters 22, 27-31.

Don, F.J.H. and J.R.Magnus, 1980, On the unbiasedness of iterated GLS estimators, Communications in statistics - theory methods A9.

Galbraith, R.F. and J.I. Galbraith, 1974, On the Inverse of Some Patterned Matrices Arising in the Theory of Stationary Time Series, Journal of Applied Probability 11, 63-71.

Gooijer, J.G. de, 1978, On the Inverse of the Autocovariance Matrix for a General Mixed Autoregressive Moving Average Process, Statistische Hefte 19/2, 114-123.

Greenstadt, J, 1967, On the relative Efficiences of Gradient Methods, Mathematics of Computation, 21, 360-367.

Harvey, A.C., 1981, Times Series Models, Philip Allan Publishers, Oxford.

Judge, G.C., W.E. Griffiths, R. Carter Hill, H. Lütkepohl and Tsoung-Choa Lee, 1985, The Theory and Practice of Econometrics, Wiley, New York.

Kennedy, W.J. and J.E. Gentle, 1980, Statistical Computing, Marcel Dekker, Inc., New York

Knottnerus, P., 1989, Linear models with correlated disturbances, Ph.D. thesis, University of Amsterdam, Amsterdam

Kohn, R. and C.F. Ansley, 1985, Computing the Likelihood and its Derivatives for a Gaussian ARMA Model, J. Statistical Computation and Simulation, 22.

Leeuw, J.L. van der, 1994, The covariance matrix of ARMA errors in closed form, Journal of Econometrics, 63, 397-405.

Magnus, J.R., 1978, Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix, Journal of Econometrics, 7, 281-312.

Magnus, J.R. and H. Neudecker, Matrix Differential Calculus, Wiley, Chichester, 1988.

Malinvaud, E, 1970, Statistical methods of econometrics, 2nd ed., North-Holland Publishing Company, Amsterdam.

McLeod, I., 1975, Derivation of the Theoretical Autocovariance Function of

Autoregressive-Moving Average Time Series, Applied Statistics 24, 255-256.
McLeod, I., 1977, Correction to Derivation of the Theoretical Autocovariance Function of Autoregressive-Moving Average Time Series, Applied Statistics 26, 194.

Nerlove M., D.M. Grether and J.L. Carvalho, 1979, Analysis of Economic Time Series, Academic Press, New York.

Newbold, Paul, Christos Agiakloglou and John Miller, 1994, Adventures with ARIMA software, International Journal of Forcasting, 10, 573-581.

Oberhofer, W. and J. Kmenta, 1974, A general procedure for obtaining maximum likelihood estimates in generalized regression models, Econometrica, 42, 579-590.

Pagan, A, 1974, A Generalised Approach to the Treatment of Autocorrelation,
Australian Economic Papers 13, 267-280.

Rao, C.R. 1973, Linear Statistical Inference and Its Applications, Wiley, New York.

Tiao, G.C. and M.M. Ali, 1971, Analysis of Correlated Random Effects: Linear Model with Two Random Components, Biometrika, 58, 37-51.

Tunnicliffe Wilson, G., 1979, Some Efficient Computational Procedures for High Order ARMA Models, Journal of Statistical Computation and Simulation 8, 301-309.

Zinde-Walsh, V., 1988, Some exact formulae for autoregressive moving average processes, Econometric Theory, 4, 384-402.

Zinde-Walsh, V., 1990, Errata (to Some exact formulae for autoregressive moving average processes), Econometric Theory, 6, 293.

Zinde-Walsh, V. and J.W. Galbraith, 1991, Estimation of a Linear Regression Model with Stationary ARMA(p,q) Errors, Journal of Econometrics, 47.

AUTHOR INDEX

SUBJECT INDEX

In 1982, T.W. Anderson wrote in a survey article on maximum likelihood estimation in autoregressive and moving average models: *A detailed study of the likelihood function ... will involve explicit expressions for the determinant and the components of the inverse of the covariance matrix. These are not available in the literature in closed form expressions for arbitrary finite orders.*

This study presents such a closed form and shows how profitable it is. Starting point is the formulation of the well known ARMA equation in matrix form by using simple triangular matrices containing the parameters. The closed form of the ARMA covariance matrix is expressed as a function of these matrices and is surprisingly simple. Its determinant and its inverse can easily be obtained, just like first and higher order derivatives with respect to its parameters.
Detailed descriptions of efficient algorithms for the covariance matrix itself, its inverse and determinant are included. Of course, pure AR and MA models are contained in the general approach as special cases.

Maximum likelihood estimation of exact ARMA models is investigated thoroughly, including first and second order derivatives of the likelihood function. A comparison of the maximum likelihood approach and the commonly used conditional least squares method for estimating ARMA parameters concludes this book.

Tilburg University Press
P.O. Box 90153
5000 LE Tilburg
The Netherlands

ISBN 90-361-9867-4